Examining Cultural Representation in Generative Image AI Models: A case study of Stable Diffusion

Master's thesis for obtaining the academic degree

Masters of Science

in "Transition, Innovation and Sustainability Environments" (TISE) programme

submitted by

Ngan Pham

Department for Knowledge and Communication Management

at University for Continuing Education Krems

Advisor: Susan Levvy

Advisor: Daniel Cardoso

Austria, 2023

STATUTORY DECLARATION

I, Ngan Pham, hereby declare,

1. that I have written my Master's thesis myself, have not used other sources than the ones stated and moreover have not used any illegal tools or unfair means,

2. that I have not used my Master's thesis or parts thereof as an exam paper in my domestic or any foreign country in any form to this date,

3. that, in case my Master's thesis concerns my employer or any other external cooperation partner, I have fully informed them about title, form and content of the Master's thesis and have his/her permission to include the data and information in my written work.

ABSTRACT

This study investigates AI fairness in cultural representation, focusing specifically on generative Text-to-Image AI Models, an emerging research area. Through a case study on Stable Diffusion, the research examines cultural representation by analyzing a diverse group of 7 cultures. Employing a mixed-methods approach involving quantitative image quality assessment, computer vision detection, surveys, and socio-semiotic analysis, the study uncovers both strengths and limitations in cultural representation.

The findings highlight instances of inaccuracies and biases across all cultures, revealing the presence of more negative biases in images representing Oriental cultures in the global South. Additionally, the study identifies gender disparities and cultural biases within and across cultures, shedding light on underlying gender inequalities. By contextualizing these biases within broader societal frameworks, incorporating concepts such as Orientalism, colonialism, and intersectionality, the study explores the causes and far-reaching implications of these biases, particularly in perpetuating stereotypes and inequalities against marginalized cultures. The research also proposes potential mitigation strategies and underscores the need for improvements in AI-generated image assessment methods.

This research contributes to the field of AI fairness by uncovering multifaceted cultural biases within generative AI image models and enhancing the shared understanding of the complex socio-techno dynamics of cultural representations in AI. The findings underscore the critical importance of examining cultural biases and striving for accurate and fair cultural representation in AI technologies.

Keywords:

cultural representation, artificial intelligence, generative AI, text-to-image AI, Stable Diffusion, biases, responsible AI, human-machine interactions.

CONTENTS

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	**Definition**
AI	Artificial Intelligence
T2I	Text-to-Image
SD	Standard Deviation
BRISQUE	Blind/ Referenceless Image Spatial Quality Evaluator
NIMA	Neural Image Assessment
MEDC	More Economically Developed Countries
CNNs	Convolutional Neural Networks
AEs	Autoencoders
GANs	Generative Adversarial Networks
VAE	Variational Autoencoders
DPM	Diffusion Probabilistic Model
NCE	Noise-Contrastive Estimation
CCUB	Cross-Cultural University of Bristol Scale
DDIM	Dynamic Dimensionality Identification Method
IQA	Image Quality Assessment
NR-IQA	Non-Reference Image Quality Assessment Metric

Introduction

Background and Motivation

The field of artificial intelligence (AI) has made significant progress in recent years, revolutionizing various fields, and transforming how humans interact with technology. Generative image AI models, such as Stable Diffusion, have gained considerable attention for their ability to generate realistic and diverse images. With the rapid advancement and integration of generative AI technologies into various aspects of our lives, it is essential to examine their applied algorithms and real-life applications and continually build ethical AI models that align with humanity's values. While these generative AI models have shown remarkable success in multi-modal generation tasks, there are growing concerns about negative biases in these models. For example, an article called “Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings” by Bolukbasi et al. (2016) examines the problem of bias in word embeddings, showing that the AI text model associates certain professions or attributes with certain genders, reinforcing gender stereotypes and discrimination.

To address these issues, various approaches have been proposed, such as using diverse datasets and training models with an explicitly defined objective function to reduce bias in gender (Salminen et al., 2020). Despite these efforts, existing studies have tended to focus on biases and representations of racial and gender minorities. There remains a research gap in examining bias in generative AI models regarding cultural representation. The lack of research on national representation in AI goes hand in hand with the imbalance distribution of AI power. As the so-called AI race heats up, research has shown that across all domains, the development and application of AI are heavily concentrated in a handful of big states and global tech giants (Mladić, 2021; Righi et al., 2020). For example, the U.S. dominates the global AI talent pool, hosting nearly 60 percent of the world’s top-tier researchers (The Global AI Talent Tracker, 2020). Looking specifically into the field of AI ethics, the power dynamic is similarly skewed. Western developed countries (MEDC), such as the U.S. and the U.K. together account for over one-third of all AI ethics principal documents (Jobin et al., 2019). Taken altogether, these imbalances raise concerns over which entities have the techno-socio-economic power to not only develop but also decide what is considered an accurate and fair representation of cultures in AI.

Cultural representation refers to the incorporation of accurate and fair cultural perspectives, identities, and experiences within AI systems. Biases in cultural representation in AI models, on the other hand, poses significant challenges, including reinforcing biases, perpetuating stereotypes, and marginalizing certain cultures and communities. Additionally, AI's influence extends to domains such as advertising, education, and healthcare, where cultural misrepresentation can lead to inequitable outcomes and limited opportunities for certain groups. Thus, the consequences of misrepresentation of culture can range from exclusionary and homogenous media content to perpetuating cultural stereotypes and exacerbating discriminatory prejudices.

This research aims to investigate the cultural representation of different cultures, specifically between dominant and less dominant cultures, in the context of Stable Diffusion, a prominent generative image AI model. By examining existing literature, and conducting a case study on Stable Diffusion, this thesis seeks to explore the model's capacity to generate images that accurately represent different cultures and assess any potential cultural biases or underrepresentation. Through this examination, the study aims to shed light on the broader challenges and opportunities associated with cultural representation in generative AI models and contribute to the ongoing discourse surrounding the ethical development and deployment of AI technologies.

Research question

Conducting a case study on a state-of-the-art Generative AI image model, Stable Diffusion, this research aims to answer the following questions:

1. How are different cultures represented in Stable Diffusion regarding 1) cultural accuracy and 2) fairness?
2. Are there any cultural biases and stereotypes of cultures reflected in Stable Diffusion's output, especially those from non-Western cultures in the Global South?
3. What are the implications of these cultural representations?

Significance and potential impacts

While addressing gender and race bias in generative AI is crucial, it does not adequately cover the breadth of bias, especially those related to the economic imbalances that exist in the world. Historically, Western countries have dominated the development of AI technology, resulting in an over-representation of Western cultures and identities in technology in general and AI-generated images. This perpetuates the marginalization and prejudices against certain groups of people from disadvantaged cultures and countries. This is especially important when considering the impact of AI-generated images in various fields, within and across global cultures, such as global recruitment, marketing, entertainment, and criminal justice. These images can also have significant consequences on how people perceive themselves and others, and the perpetuation of cultural stereotypes can have a lasting impact on the social and political landscape.

Zooming out, the significance of addressing cultural representation in AI lies in its potential to shape societal dynamics and influence human-machine interactions. In the context of modern-day society, the misrepresentation of certain historically underprivileged or marginalized cultures would undermine the empowerment potential of AI and instead reinforce historical power imbalances, between the developed and underdeveloped world, and between North and South. To ensure ethical AI is ethical globally and across cultures, thus, is to examine and mitigate potential cultural biases.

In conclusion, it is essential to examine cultural representations in generative AI models to ensure that these models accurately and fairly reflect the diverse cultures of our global society. By doing so, we can work towards a more inclusive and equitable development of AI technology, one that reflects the diversity of human identities, and empower those from historically disadvantaged cultures to get equal benefits from these emerging technologies.

Literature review

Generative Text-to-Image (T2I) Model

Generative Image AI Overview

Generative AI models have been a topic of interest in the field of computer science for several years (Song et al., 2020). These models can generate new data that is like the training data, which has led to their use in various applications such as text-to-speech, image-to-image translation, and text-to-image generation. These generative AI image models have shown potential applications in bridging the gap between language and visual representations and have been an active research area in the last decade (Zhang et al., 2023). Beyond direct applications in tasks such as image creation and editing in visual arts and computer graphics, more and more scientists are also exploring the use of generative image models in healthcare, for example, creating synthetic data for medical training, building decision support systems and protecting patient privacy (Jadon & Kumar, 2023).

Generative Text-To-Image (T2I) Models

In the context of Text-to-Image (T2I) generation, these generative AI models are a type of generative model that can generate images based on free-form and open-ended textual input (Qadri et al., 2023). These models employ different types of deep learning methods such as convolutional neural networks (CNNs), auto-encoders (AEs), and generative adversarial networks (GANs) to generate images based on a given prompt (Intodia et al., 2023). Recent breakthroughs in generative AI image models include popular models such as DALL-E and Stable Diffusion, which have become state-of-the-art in many computer vision and image processing problems (Zhang et al., 2023).

One popular model is a generative adversarial network (GAN), introduced by Goodfellow et al. (2014) and revolutionized the field of generative AI models. GANs consist of two neural networks: a generator and a discriminator. The generator creates images from noise, while the discriminator tries to distinguish the generated images from real images. The generator and discriminator are trained simultaneously until the generated images are indistinguishable from real images (Frolov et al., 2021). In the case of Text-To-Image models, the generator takes in the feature vector generated by the text encoder and produces an image that is intended to match the textual description. The discriminator is then trained on a dataset of real images and images generated by the generator. The discriminator's goal is to correctly identify whether an image is real or generated. The generator is trained to create images that are increasingly difficult for the discriminator to distinguish from real images.

Another type of generative model is Variational Autoencoders (VAE), proposed by Kingma and Welling (2013). VAE models typically consist of two main components: a text encoder and an image decoder. The text encoder is a neural network that takes in textual input and generates a feature vector representation. The image decoder takes this feature vector as input and generates an image that is intended to match the textual description. They are similar to GANs but use a different approach to generate images. VAEs can generate more diverse images than CGANs and can handle missing data better. However, they can also suffer from mode collapse and require a large amount of training data.

Diffusion models have recently gained attention as a promising approach for generative image modelling. Inspired by the theory of partial differential equations, diffusion models employ a diffusion process to generate images. Ho et al. (Ho et al., 2020) introduced the Diffusion Probabilistic Model (DPM), which formulates image generation as a diffusion process with progressively increasing noise levels. Later, Sohl-Dickstein et al. (2021) proposed the Noise-Contrastive Estimation (NCE) framework, which enables training diffusion models using noise-contrastive objectives. Unlike GANs, which use a generator and discriminator, diffusion models use a diffusion process to gradually change noise into an image. Diffusion models are trained to optimize a likelihood function that estimates the probability of the generated image given the input text.

Diffusion models have demonstrated impressive results in generating high-resolution images with fine-grained details. One advantage of diffusion models over GANs is that they can generate high-resolution images with much higher fidelity than GANs. Additionally, diffusion models can generate diverse images that are not similar to each other, unlike GANs. Another notable study by Song et al. (2020) built upon the concept of diffusion generative models and proposed Diffusion Probabilistic Models (DPMs) for image generation. The authors showed that DPMs can achieve state-of-the-art performance in image generation while maintaining a simple and interpretable model structure. More and more studies have continued building on diffusion models and enhancing their performance in various applications. For example, other researchers have explored the use of diffusion image models in image inpainting (Xie et al., 2021), image super-resolution (Gao et al., 2020), and image compression (Li et al., 2021).

All in all, diffusion models are state-of-the-art generative image AI models that hold great promise for generating high-quality images for a variety of applications. The development of recent models has resulted in the emergence of several innovative applications beyond the creation of images from text, including artistic painting and text-guided editing (Zhang et al., 2023). As more research is carried out, we can expect to see even more sophisticated and versatile approaches to image generation that take advantage of the unique capabilities of diffusion processes.

AI Applications and Implications for Fairness and Diversity

With its increasing power and popularity, generative image AI has numerous potential real-world applications across various fields. For example, Generative models such as GANs, VAE, and GSN can be used to translate images from one setting to another while keeping the structure of generated images aligned with the input images. This can be useful in automating web design, where generative models can be used to automatically generate colorization for web pages based on user-defined parameters (Kikuchi et al., 2023).

Specifically, conditional generative models such as DALL-E and Stable Diffusion can be useful in various applications, such as generating images for arts and game design, where generative models can be used to generate images based on user-defined prompts (Deckers et al., 2023). AI models have also been used in fashion and design to generate images of clothing items, which can be used to create virtual try-on systems or to generate new designs. In 2017, for example, a team of researchers used a GAN to generate images of shoes, which were then used to train a classifier that could recognize different shoe styles. In media and entertainment, Generative models can be used to generate realistic animations for movies and TV shows. For example, generative models can be used to generate realistic facial expressions and movements for animated characters.

With its ever-growing popularity and widespread adaptation for both research and personal and commercial applications, there is growing interest in analyzing its implications for real-world applications. One important topic is to ensure that the AI model and the images generated are appropriate, ethically used, and equitable for people of different sociodemographic groups, such as race, gender, and culture. While generative image AI models offer numerous benefits, they also raise ethical concerns and implications. One major concern is the potential for generating deep fakes, where malicious actors can manipulate images and videos to deceive or defame individuals. This poses challenges for trust, privacy, and digital forensics. Moreover, biases in training data can lead to biased outputs, perpetuating societal inequalities and reinforcing existing prejudices. Ensuring fairness, transparency, and accountability in generative image AI models is crucial to mitigate these ethical risks.

Generative AI Images and Cultural Diversity

The question of cultural diversity and representation within generative AI images has gained significant attention in recent years. Ensuring cultural diversity and fairness in AI is a challenging endeavor as studies have shown that AI exhibits a wide range of systemic biases that are not fully understood.

One issue identified is bias in the training data set. Generative AI models heavily rely on large-scale datasets for training. However, these datasets often suffer from biases and underrepresentation of certain cultural groups. For example, studies have found that datasets used to train generative AI models predominantly feature images of individuals that led to the inconsistencies and biases of machine learning algorithms in identifying the gender and skin type of certain minority groups, especially “darker females” (Buolamwini and Gebru, 2018). Similarly, this lack of diversity can result in generated images that do not accurately represent or adequately reflect the visual characteristics, cultural nuances, and diversity of different cultures, especially those of marginalized groups and traditionally underrepresented.

Correspondingly, the biases present in training data can be perpetuated and amplified in the generated images produced by generative AI models. These biases can manifest in various forms, such as skin tone disparities, stereotypical depictions, or the exclusion of underrepresented cultural identities. For instance, Geirhos et al. (2020) found that popular object recognition models were biased towards lighter skin tones, leading to higher classification errors for darker-skinned individuals. Similar biases may be present in generative AI models, affecting the representation and portrayal of cultural diversity.

While there is a lack of specific studies on the implications of AI for cultural representation, the existing research on gender and racial bias in AI can provide insights. It is reasonable to argue that the presence of cultural bias and underrepresentation in the dataset could result in inaccurate cultural representation in generated images. Such misrepresentations can have detrimental effects on both the individuals depicted in distorted or unfair ways and the audiences consuming these representations. AI, as a new tool for media generation, has the potential to reinforce stereotypes, marginalize cultural groups, and perpetuate inequalities, similar to mainstream media (Hall, 1997).

For instance, studies have demonstrated that the media's perpetuation of "benevolent sexism," portraying women as inherently better caregivers, reinforces gender inequality and leads to workplace discrimination against women (Glick & Fiske, 1996). Correspondingly, a recent study on gender fairness in Stable Diffusion has found that the model produced more female faces in images of care-taking jobs, such as nurses, while it is the opposite for managerial roles (Friedrich et al., 2023). This indicates that the AI model, trained on web-scraped data, exhibits gender bias in occupation and perpetuates long-standing stereotypes prevalent in mainstream media. These biased representations in AI-generated images can contribute to the perpetuation of "benevolent sexism" and limit opportunities for individuals who do not conform to gender stereotypes. This creates a vicious cycle where the bias present in the training dataset is reproduced through AI generation, further perpetuating the bias and providing new biased media as data for training. The same can be extrapolated to biased cultural representations, emphasizing the importance of addressing these biases in AI models. Additionally, generative AI images that fail to capture the diverse range of cultural identities and experiences can contribute to a sense of exclusion and alienation for individuals from underrepresented communities.

To address this, researchers and practitioners are developing strategies to improve the representation and accuracy of generated images across diverse cultural backgrounds. This includes the development of more inclusive and diverse training datasets that encompass a broader range of cultural identities and characteristics (Friedler et al., 2018; F. Liu et al., 2021). Additionally, techniques such as data augmentation, domain adaptation, and fine-tuning on culturally diverse datasets can help mitigate biases and enhance cultural diversity in generated images. Data augmentation involves generating additional training data by applying transforms to an existing dataset, while domain adaptation involves adapting a model trained on one dataset to another, related dataset, which can help to improve the models' performance across diverse cultural contexts. Finally, fine-tuning involves training a model an existing pre-trained model, on a new dataset, which can help to further adapt the model to a specific task, including diverse cultural understandings (Friedler et al., 2018). For example, a recent study proposed The Cross-Cultural Understanding Benchmark (CCUB) Dataset that consists of culturally-diverse training data that can be used to fine-tune AI models and improve their ability to understand and respond to diverse cultural contexts (Z. Liu et al., 2023). From a human-centric approach, engaging users from diverse cultural backgrounds in the generative AI image-creation process can also contribute to more culturally sensitive and inclusive outcomes (Qadri et al., 2023).

Cultural representation

Defining Cultural Representation

The first aspect of culture and cultural representation is how culture is defined and understood. Culture is a complex and multifaceted concept that has been approached from a variety of theoretical perspectives. A dominant definition of culture comes from cultural studies, which define culture as a set of shared beliefs, values, practices, and artifacts that define a particular group or society (Shome, 2019). Culture shapes and is shaped by social, political, and economic forces. Key themes that the studies of culture explore include Identity and difference, globalization, and power and representation. In the context of ethical AI, the latter is particularly relevant. For decades, there have been different theorists that proposed and analyzed how representations and power relations are intertwined. Marxist Antonio Gramsci argued that power is exercised and contested through cultural practices and representations, and thus representation plays a key role in how the ruling class portrays itself and others to maintain its power over society, including ideas about race, culture, and gender. Meanwhile, in Michel Foucault's theory of discourse, representation is viewed as a means of exerting power and representations construct meaning and shape our understanding of the world (Stoddart, 2007). While the two theorists differ in some of their key ideas, they share an interest in representation's role in shaping power relations.

Taking about culture specifically, cultural representation is a complex concept that embodies a range of meanings and interpretations. From an etymological perspective, representation can be interpreted as a creation not by depicting objects in their original form, but rather by constructing a new form or context. Initially, in ancient times, representation served as a pivotal element of literature, aesthetics, and semiotics. Nowadays, it has become a significant component of the contemporary art landscape encompassing audio, visual, and textual mediums like movies, painting, literature, and advertisements.

One of the pioneering works on representation is sociologist Stuart Hall’s The Work of Representation (Hall, 1997). According to Hall, representation refers to the process by which meaning is produced and exchanged within a culture. This process involves the use of signs and symbols, which are invested with meaning by the culture in which they are produced and received. These signs and symbols are not neutral but rather are shaped by the cultural values and ideologies that inform their use. Hall argues that representation is a complex and contested process, as it involves the negotiation and contestation of meanings within and across different cultural groups. Different groups may have different interpretations of the signs and symbols used in cultural texts, and these interpretations may be influenced by factors such as social class, race, gender, and sexuality. In this way, representation is not a fixed or stable concept but rather is subject to change and transformation as cultural values and ideologies evolve over time. Hall's work emphasizes the importance of critically examining how meaning is produced and exchanged within cultural texts, and of recognizing how representation is shaped by power relations and social hierarchies.

Importantly, Hall’s work emphasizes the importance of cultural representation and how language and knowledge production systems work together to create and perpetuate these meanings. According to him, people assign meaning to things through their representation of them, and cultural representations help shape the way individuals view themselves and others within society. These representations can also become institutionalized and influence policies and practices. Hall emphasizes the role of power in the politics of representation, as individuals and groups seek to control the representation of themselves and their culture, while also facing the potential for misrepresentation by others.

In the field of cultural studies, representation is used to critique and challenge existing boundaries between minority and dominant cultures, between disciplines, between academic and street cultures, between ethnicities, races, genders, and classes, between student texts, literary texts, and other texts, and between representation and the represented (Pari, 1992). It is concerned with how cultural objects and practices are constructed and disseminated through various media, and how representations of race, gender, class, and other social categories are circulated and contested in popular culture. Another definition comes from Cultural semiotics, which studies the symbols or the representation system used by human beings to give things meaning (Long & He, 2021). Cultural representation in cultural semiotics refers to how cultural objects, practices, and phenomena are constructed and communicated through signs and symbols within a particular cultural context.

In both fields, the study of cultural representation involves questioning how cultural identities and meanings are constructed and communicated through representation, but the emphasis and method of analysis differ. Cultural semiotics looks at how signs and symbols are used to construct meaning, while cultural studies explore how cultural representations are shaped through power relations and social categories. This project focuses on the latter, i.e. examining the cross-cultural representations of under-represented groups in generative T2I AI models, and thus, primarily adopting the cultural representation definition in cultural studies.

Cultural Representations and Power Dynamics

The relationship between cultural representations and power dynamics has been a key topic of inquiry in academic literature across a range of disciplines, including cultural studies, media studies, and sociology. Scholars have explored how cultural representations can shape and reflect power imbalances in society. For example, people of color are often portrayed in ways that reinforce racial stereotypes and marginalize them within mainstream culture (Hall, 1997). These representations can contribute to the oppression and marginalization of these groups, and perpetuate power imbalances.

Another important dimension of this relationship is the way that cultural products are produced and consumed. Cultural industries such as Hollywood and the music industry are often dominated by a small number of powerful corporations and individuals, who can shape cultural representations in ways that reflect their interests (McChesney, 2008). This can contribute to the marginalization of certain groups and the reinforcement of power imbalances.

At the same time, cultural representations can also be used to challenge power imbalances and promote social justice. For example, "Black Looks: Race and Representation" by bell hooks explores the personal and political consequences of contemporary representations of race and ethnicity (hooks, 1992). hooks interrogated old narratives and argue for alternative ways to look at Blackness and the importance of challenging dominant representations that perpetuate racism and cultural stereotypes, calling for a shift towards more diverse and inclusive representations that reflect the complexities and diversity of Black experiences. In parallel, critical media scholars have highlighted the importance of media products such as literature, film, and music in promoting the experiences and perspectives of marginalized groups (Kellner & Share, 2007). This effect can be extended to cultural representations in media. In other words, when marginalized groups can create and disseminate their cultural representations, they can use these representations to challenge dominant narratives and assert their own identities and experiences. Furthermore, research has shown that exposure to positive cultural representations of marginalized groups can lead to more positive attitudes and reduced prejudice toward those groups. For example, a 2022 literature review found that exposure to TV shows with a positive, diverse representation of ethnic minorities can lead to more positive attitudes toward marginalized groups, including increased empathy and reduced prejudice (Żerebecki et al., 2021).

In conclusion, the literature on cultural representations and power dynamics highlights the complex and multifaceted relationship between cultural products and social inequality. While cultural representations can both reflect and reinforce power imbalances, they can also be used to challenge dominant narratives and promote social justice. Considering the increasing importance of diversity inclusion and emerging technologies, further research is needed to better understand the mechanisms through which cultural representations in media, particularly in new AI-enabled media such as AI-generated images, impact the power dynamics today.

Cultural Stereotypes

Defining Cultural stereotypes

Cultural stereotypes refer to widely held beliefs, assumptions, or generalizations about individuals or groups based on their cultural background. These stereotypes can shape perceptions, attitudes, and behaviors toward specific cultures, often leading to oversimplifications, biases, and misunderstandings. In this literature review, we delve into the concept of cultural stereotypes, examining their definitions, formation processes, and implications within various disciplines.

Scholars from different fields have provided definitions and conceptualizations of cultural stereotypes. Most notably, Hofstede (1991) defines cultural stereotypes as "shared beliefs about the characteristics, attributes, and behaviors of members of a specific culture." These stereotypes are often based on limited or incomplete information and can lead to assumptions about cultural groups. Other researchers emphasize that cultural stereotypes involve the attribution of specific traits, values, or behaviors to individuals solely based on their cultural affiliation (Lee et al., 2000). These definitions highlight the cognitive processes underlying the formation of cultural stereotypes.

Cultural stereotypes are shaped by various factors, including media representations, socialization processes, personal experiences, and intergroup dynamics. Media plays a crucial role in perpetuating cultural stereotypes by disseminating images, narratives, and portrayals that reinforce preconceived notions about different cultures (Entman, 1994). Socialization processes, such as family, education, and peer interactions, also contribute to the formation and maintenance of cultural stereotypes (Rubin and Hewstone, 1998). Furthermore, intergroup dynamics, such as ingroup-outgroup biases and intercultural contact, influence the development and persistence of stereotypes (Pettigrew and Tropp, 2006).

While cultural stereotypes can sometimes contain elements of truth, they often oversimplify and exaggerate cultural characteristics. Research has highlighted the limitations and inaccuracies of cultural stereotypes, emphasizing that they do not capture the full complexity and diversity within cultural groups (Judd and Park, 1993). Cultural stereotypes tend to overlook individual variations, context-specific behaviors, and changes over time, leading to biased and distorted perceptions of cultural groups.

Cultural stereotypes have significant implications for intergroup relations, social interactions, and societal outcomes. Stereotypes can lead to prejudice, discrimination, and unequal treatment of individuals based on their cultural background (Fiske et al., 2002). Stereotype threat, a phenomenon where individuals feel pressured to conform to negative stereotypes, can hinder performance and well-being (Steele and Aronson, 1995). Moreover, cultural stereotypes can perpetuate inequalities and hinder cross-cultural understanding and collaboration.

Over the last decades, with growing globalization and increased access to media and travel, efforts have been made to challenge and debunk cultural stereotypes through education, intercultural training, and exposure to diverse cultural perspectives (Paolini et al., 2010). Promoting intercultural competence, which involves developing knowledge, attitudes, and skills for effective intercultural interactions, can help reduce the influence of cultural stereotypes and foster greater cultural understanding and empathy (Deardorff, 2006). Furthermore, given the changing dynamics of our society with immigration, use of social media and technology, interdisciplinary collaborations must be made to shed light on the influence of cultural stereotypes across various domains and settings, and work across fields such as education, media, and technology, to collaboratively work towards combating negative stereotypes and build towards more cultural diversity in our embedded socio-technological systems.

Cross-Culture Studies: East Versus West

One of the seminal works in cross-cultural comparison and defining cultural realms is Hofstede's book, "Culture's Consequences: International Differences in Work-Related Values" (Hofstede, 1983). In this book, Hofstede identifies several cultural dimensions that differentiate various societies. Based on his survey of more than 116,000 IBM employees in 40 countries, the team came up with Hofstede's Cultural Dimensions Theory as a framework for capturing and comparing cultural values across cultures using six dimensions: Power Distance, Uncertainty Avoidance, Individualism/Collectivism, Masculinity/Femininity, Long/ Short Term Orientation, and Indulgence/Restraint.

While there is no definite category of "Western" or "non-Western" culture, Hofstede's cultural dimensions offer a framework to gauge cultural proximity between 2 dominant culture spheres from the West and non-Western countries. In particular, some of his dimensions have been associated with cultural values that are commonly associated with Western cultures. For example, the dimension of individualism-collectivism refers to the degree to which individuals prioritize their interests versus the interests of their group. Western cultures, such as those in North America and Western Europe, tend to be more individualistic, while non-Western cultures, such as those in East Asia and the Middle East, tend to be more collectivistic. The dimension of masculinity-femininity refers to the degree to which a culture values assertiveness, competitiveness, and achievement (masculine) versus nurturing, cooperation, and quality of life (feminine). Some researchers have suggested that Western cultures, particularly those in North America and Western Europe, tend to be more masculine than non-Western cultures.

In general, Western cultures tend to score higher in individualism, and masculinity, and lower in power distance and uncertainty avoidance compared to non-Western cultures. They also tend to score lower in long-term orientation compared to some non-Western cultures. However, it is important to note these trends are not absolute and can vary significantly within and across Western cultures. Cultural values and practices are complex and multifaceted and the dimensions should be used solely for broad generalizations.

Orientalism

A concept closely related to cross-cultural studies and cultural stereotypes is “Orientalism”. First coined by the literary critic Edward Said in his book "Orientalism", it is a term used to describe how Western societies historically viewed and represented the cultures and peoples of the Middle East, Asia, and North Africa and refers to a set of stereotypes, assumptions, and generalizations about the Orient as the “Other” (Said, 1978). Said argues that Orientalism is a form of “cultural imperialism” that is based on the assumption that the West is superior to the East in terms of culture, civilization, and values. For Said, Orientalism is the product of two centuries of Western scholarship and of British, French, and more recently American political and social needs for domination. It involves the representation of the Orient as exotic, primitive, and irrational, and portrays Western culture as rational, progressive, and superior.

In other words, orientalism is a form of cultural representation that has contributed to negative attitudes and perceptions towards the cultures and peoples of the East from the West. For example, it has been used as a tool to create demeaning, degrading, reductive descriptions of the people of the Arabs and Islam (Said, 1978). Orientalism has also had a long-lasting impact on the representation of Asian women in popular media and cultural artifacts, where it has underscored reductive and damaging stereotypes of hyper-sexualization and objectification (Matsumoto, 2020).

Orientalism is a complex concept that has generated fierce and unrelenting debates regarding its epistemology and scope and has undergone changes in various ways from Said’s times. Modern scholars argue orientalism has mutated to engender a paradigmatic change, and a more sophisticated, subtle, and up-to-date perspective has appeared as a result of the rise of global capitalism, cultural homogenization, and digital technology (Hodkinson et al., 2013). Although its emphases, concerns, and methodologies might represent a certain departure from old orientalist dogmas, its objective seems to remain largely intact. The scholars also highlight the multiple forms of orientalism within the "West," the manifold presence of the "East" in the Western world, and the epistemological fragility of the ideas of "Occident" and "Orient" as such.

While Orientalism is rooted in colonialism, these power dynamics influenced the formation of cultural stereotypes that continue to shape perceptions of diverse cultures today. The economic inequalities resulting from colonial legacies further exacerbate the misrepresentation and marginalization of cultures from the Global South, perpetuating cultural stereotypes. A recent paper named "Orientalism in a globalized world: Said in the twenty-first century" engages with these debates, exploring the continued relevance and usefulness of Said's ideas in today’s society (Sa’di, 2021). The digital age has made it easier to disseminate Orientalist ideas through various media platforms, including social media, which can spread misinformation and stereotypes about the "Orient". Further, Sa’di argues that while Said's ideas are still relevant in the contemporary world, postcolonial scholars need to take account of new forms of power and inequality that have emerged and adapt a nuanced and sensitive approach to questions of identity and belonging that extends beyond Said's original analysis.

AI systems are not immune to the biases and cultural stereotypes that emerge from Orientalist discourses. The predominance of Western cultural perspectives and datasets in AI training exacerbates the representation gap, leading to distorted and oversimplified depictions of non-Western cultures. This perpetuates power imbalances, reinforces stereotypes, and marginalizes cultures from the Global South. As such, examining and understanding embedded Orientalism in the AI model allows for a nuanced examination of power imbalances and the historical legacies of colonialism in emerging and subtle formats that are both relevant and hidden in today's society. To advance cultural diversity in AI and combat stereotypes, it is essential to examine the impact of Orientalism within the broader context of colonialism and economic inequality. By recognizing and challenging the dominant Western narratives, researchers can advocate for more equitable representation and participation of diverse cultural groups. Such efforts can foster cultural diversity, empower marginalized voices, and promote a more equitable AI landscape.

Intersectionality

Intersectionality, a framework developed by Crenshaw (1989), highlights the interconnected nature of social identities and the overlapping systems of discrimination and privilege. Intersectionality emphasizes how various dimensions of identity, such as race, gender, class, and sexuality, intersect and interact to shape individuals' experiences and social inequalities. In the context of combatting cultural stereotypes, understanding and applying intersectionality can provide valuable insights into the complex dynamics of cultural representation and challenge the oversimplified narratives associated with stereotypes such as Orientalism. By considering the intersection of race, ethnicity, gender, religion, and other relevant aspects, researchers can gain a deeper understanding of the diverse ways in which Orientalism stereotypes manifest and impact individuals from different cultural backgrounds.

Intersectionality calls for more authentic and diverse representation that goes beyond simplistic and essentialist portrayals. By incorporating intersectional perspectives, cultural representations can challenge Orientalism stereotypes and provide counter-narratives that highlight the complexities, agency, and diversity within Asian and Middle Eastern cultures. This involves giving voice to individuals from these cultures, allowing for self-representation, and actively involving them in the creative process. Authentic representation resists Orientalism stereotypes and promotes a more accurate and respectful understanding of cultural diversity. Correspondingly, cultural representation in AI should adopt an intersectional lens to acknowledge and address the complex ways in which multiple cultural identities intersect and shape individuals' experiences (Noble, 2018). Failing to consider intersectionality can lead to incomplete or biased representations of cultural diversity.

Cultural Representation in AI and Ethical Considerations

Taking into account perspectives from sociology, media, and information system studies, cultural representation in AI involves capturing and incorporating cultural nuances, symbols, and meanings to ensure accurate and respectful depictions of various cultures. Cultural representation in AI, thus, is rooted in the broader societal recognition of the importance of diversity, equity, and inclusion. Similar to media, scholars argue that AI systems should not perpetuate biased and exclusionary narratives but should strive to accurately reflect the rich diversity of cultures present in society (Holstein et al., 2019).

Achieving cultural representation in AI is not only an ethical imperative but also a means to challenge and dismantle existing power structures and biases (Ruha Benjamin, 2021). Furthermore, cultural representation aligns with principles of fairness and social justice, as it ensures equal opportunities and rights for all individuals irrespective of their cultural backgrounds.

In short, cultural representation in AI entails the inclusion and accurate portrayal of diverse cultural perspectives and experiences. It is an ethical imperative and an important step toward building fair and inclusive AI systems. Achieving cultural representation requires addressing biases, considering cultural context, employing appropriate data collection and labeling practices, adopting an intersectional approach, and upholding ethical considerations.

Cultural Biases in AI

Bias in AI

As AI becomes more prominent in our daily lives, its deployment has created a lot of debate around ethical considerations, especially around inclusivity and fairness. AI systems, including image recognition and natural language processing algorithms, have been found to exhibit a wide range of biases that reflect stereotypes and prejudices (Buolamwini & Gebru, 2018). For example, facial recognition algorithms have shown higher error rates for individuals with darker skin tones and women, indicating systemic biases that can lead to discriminatory outcomes (Gebru et al., 2018). Similarly, natural language processing models have displayed biases in language generation, favoring certain cultural perspectives or reinforcing gender stereotypes (Bolukbasi et al., 2016). A recent paper titled "Hey ASR System! Why Aren't You More Inclusive? Automatic Speech Recognition Systems' Bias and Proposed Bias Mitigation Techniques" addresses Automatic Speech Recognition systems' bias that excludes people from various minority spectrums, namely race, gender, sick and disability (Ngueajio & Washington, 2022). Generative AI with its exponential popularity and adoption in various fields, means that it is ever more important to analyze and ensure its ethical application.

Recent studies have highlighted the issue of bias in generative image models and the need to address these biases to ensure fair and equitable outcomes. The biases in generative image models can be broadly categorized into several categories. One of the most common biases in generative image models is gender and racial bias. Text-to-image AI models may be biased towards certain gender or racial characteristics, such as generating images of people with lighter skin tones or stereotypical gender roles. For example, in the research conducted by

Buolamwini and Gebru, we see that the facial recognition systems were highly inaccurate (more than 30%) when it comes to classifying the faces of women of color. In this ground-breaking paper, the authors further demonstrated that the model was most accurate for people who identified as male and of white skin tone (Buolamwini and Gebru, 2018).

Accuracy bias is a common problem among various applications utilizing machine learning. ML systems' data, algorithms, and recommendations have been found to exhibit differences in the accuracy of prediction and machine bias concerning certain subgroups, for example, gender, and racial minorities when it comes to the automatic calculation of insurance policies (Chen et al., 2019).

Another bias related to the training dataset is Western cultural biases. Text-to-image AI models may be biased towards Western cultural norms and values, such as generating images of Western-style clothing or architecture because of the lack of training data from non-Western cultures with less economic power and limited digital dataset. This can lead to a lack of diversity in the generated images and limit the representation of other cultures and traditions. Stereotypes and generalizations are also a concern in generative image models. Text-to-image AI models may make assumptions or generalizations based on the text input, such as assuming that all people from a certain culture have a particular appearance which can lead to the perpetuation of harmful stereotypes and reinforce biases. Finally, misinterpretation of cultural cues is another concern in generative image models. Text-to-image AI models may misinterpret cultural cues or symbols, such as generating images of people in inappropriate attire for a specific cultural context or generating images that misrepresent cultural traditions or practices. This can lead to cultural insensitivity and perpetuate harmful stereotypes.

Causes Of AI Bias

Bias in machine learning and AI systems is a growing concern. There are several causes of bias in these systems, both intentional and unintentional. First, Biased training data is a prominent cause of biases in AI models. Training data often reflect societal biases and imbalances, as they are collected from various sources that might have underlying biases. Biases in training data can stem from historical discrimination, social prejudices, or unequal representation of certain groups. The unfair treatment of minorities is well-documented in the data we have created over time. These historical biases trickle into the automated systems through the data. ML/AI systems trained on these data can pick the implicit biases exercised over the years, which we might not see at first glance. For example, advertisement algorithms show more high-paying technical jobs to men than women based on the training data even without being told explicitly to do so (Shrestha & Das, 2022). In the same paper, the authors systematically review over 120 pieces of literature on gender bias in ML and argue that bias is most harmful when the bias is not as readily noticeable. Especially, in the systems of online recommendations, social programs, national defense, justice systems, and policing, which implements ML/AI algorithm, when the decision made by the automated systems might be biased but there is no definite way to confirm and result in unintentional, and yet equally harmful, biases.

Algorithmic design choices can also contribute to biases in AI models. Biases can be unintentionally introduced through the design process, such as the selection of features, preprocessing steps, or the choice of evaluation metrics. For example, if a facial recognition algorithm primarily relies on specific facial features more prevalent in certain racial or ethnic groups, it can lead to higher error rates for underrepresented groups (Buolamwini & Gebru, 2018). Biases can also emerge from the optimization objectives used during training, leading to unintended discriminatory behaviors (Zhang et al., 2018).

It is important to note that biases in AI models are not isolated from societal factors. Societal biases, stereotypes, and systemic discrimination can find their way into AI systems. Pre-existing social inequalities and historical biases can manifest in the data collection process, labeling practices, and the overall development pipeline. For instance, biases in criminal justice data can lead to unfair predictions and decisions regarding criminal recidivism or sentencing (Angwin et al., 2016). Understanding the contextual biases that AI systems inherit is essential for mitigating biases effectively.

Additionally, human biases are also prevalent. If developers have gender biases, it could lead to bias in the model's outputs. The lack of diversity within AI development teams is also a contributing factor. Homogeneous teams may inadvertently overlook or fail to recognize biases that affect underrepresented groups. Diverse perspectives and experiences are crucial for identifying potential biases and developing more inclusive and fair AI systems (Mittelstadt et al., 2019).

Overall, it is obvious that the causes of bias are complex and intertwined, and that Addressing biases requires a multi-faceted approach that involves diverse data collection practices, rigorous algorithmic design, and inclusive development processes. Recently, algorithmic transparency in ML and AI systems became an area of focus when the difficulty of identifying the sources of biases leads to human control and autonomy issues. The panacea suggested by the concept of "algorithmic" transparency and accountability is limited in various contexts, as it does not address other factors such as cultural biases or preconceptions.

Bias Mitigation Methods And Frameworks

To address these biases, several attempts in both methods and frameworks have been made. Generally, mitigation frameworks include methods that detect and counter biases in different stages of the models (Feldman & Peake, 2021). First, pre-processing techniques aim to reduce biases in training data before model training. These methods include data augmentation, data re-sampling, and demographic parity approaches. Data augmentation techniques generate additional training samples to address data imbalances and increase the representation of underrepresented groups. Demographic parity approaches focus on ensuring equalized odds or predictive parity across different demographic groups (Kamishima et al., 2012).

Second, for in-processing methods, algorithmic adjustments can modify the learning algorithms to reduce biases in decision-making. These methods include equalized odds, disparate impact, and disparate mistreatment mitigation. Equalized odds adjust the decision thresholds to ensure equal false positive and false negative rates across different groups. Disparate impact approaches modify the decision rules to achieve demographic parity. Disparate mistreatment methods aim to minimize the unfair treatment of individuals based on protected attributes (Hardt et al., 2016).

Last, post-processing interventions focus on modifying the outputs of AI models to mitigate biases. These methods include calibration, rejection, and fairness-aware post-processing. Calibration adjusts the model's output probabilities to align with the desired fairness criteria. Rejection methods involve rejecting certain predictions to prevent biased outcomes. Fairness-aware post-processing techniques modify the model's predictions based on fairness constraints to achieve improved fairness (Kleinberg et al., 2016; Pleiss et al., 2017).

The effectiveness of bias mitigation methods can vary depending on the nature of the biases in the cultural context. It is essential to consider the inherent biases present in our social systems that reflect in the data and the people involved in all processes of AI models. While no single method can eliminate biases, some approaches that have shown promise in mitigating racial biases in AI systems could be implemented for cultural biases.

First, incorporating diverse and representative training data is crucial for addressing biases against minority groups (Buolamwini & Gebru, 2018). By ensuring a broad range of cultural perspectives and demographics are included in the training data, AI models can learn to recognize and understand the nuances of different cultural groups and help mitigate biases resulting from the underrepresentation of certain groups. Currently, while there are several big image datasets used for AI training, there is no dataset specially curated for cultural appropriateness. Popular large datasets, such as ImageNet, and LION-5B which Stable Diffusion was trained on, taken randomly from the internet make them highly data-driven and prone to biased human behavior. There have been efforts more recently, for example, the MaRVL dataset that created a set of cultures and languages, including Indonesian, Swahili, Tamil, Turkish, and Mandarin Chinese, comprised of diverse cultural concepts (F. Liu et al., 2021). The downside, however, is that obtaining unbiased and representative data is a complex, time-consuming, and expensive process.

Another strategy is contextual adaptation methods which adjust the AI system's behavior based on the specific context or user preferences. This approach involves incorporating customization to the system's outputs, or dynamically adapting the system's responses to align with preferable norms and values (Gao et al., 2020). This approach has shown positive results. Recently, a group of researchers has put forward a remarkable strategy with this approach to address bias in a text-to-image model called Fair Diffusion (Friedrich et al., 2023). This approach can attenuate biases in generative image models and allows for the shifting of bias in any direction based on human instructions. In the empirical study on binary gender fairness in occupation images, the team discovered Stable Diffusion model often reflects gender bias in occupation images, for example, a more male-looking person appearing in images of firefighters. To illustrate their method, the researcher instructed the model to simultaneously suppress one and reinforce another gender so that 50% of the output is dedicated to one. This method has several advantages as it does not require data filtering or additional training, and the empirical results demonstrate that it is useful for instructing generative image models on fairness. This is a promising method that could be used for cultural adaptation in AI image models to generate more equal cultural image outputs, without the need of training a model from scratch. However, unlike binary gender classification, it is more complex to quantify the classification of cultures and determine the threshold of what is more equitable. The research acknowledges that themselves in the research and encourages future attempts to adapt their approach for non-binary fairness instruction. This is a challenging and exciting avenue to look into, for example, non-binary gender diversity and cultural diversity.

Bias Detection And Evaluation Framework

To address these biases, the detection of unwanted biases in an ML/AI model is as important as the mitigation of the biases. The detection frameworks allow the create benchmarks that model designers can use to vet these ML/AI models as well as actively mitigate biases in the models. For example, in gender bias detection, Schwemmer et al.'s comprehensive detection framework showed that FRT systems like Google Cloud Vision, Amazon Rekognition, and Microsoft Azure Computer Vision, possess gender bias, which can be tested using benchmarks created from human-coded datasets (2020).

Fairness metrics and evaluation frameworks are essential for assessing the effectiveness of bias mitigation methods. Metrics such as disparate impact, equalized odds, and statistical parity difference quantify the level of bias reduction achieved by different methods. Evaluation frameworks provide guidelines and benchmarks for assessing fairness, transparency, and accountability in AI systems. These frameworks help researchers and practitioners make informed decisions about bias mitigation approaches and compare their performance (Kusner et al., 2017; Verma et al., 2018).

While the current studies have suggested several frameworks and evaluation metrics on racial and gender fairness, not a lot of attention has been made to analyzing cultural representation in AI. Furthermore, cultural representation and biases are more implicit and harder to examine because of their intersectionality with other sociodemographic factors. Thus, more research is needed to ensure fair and accurate cultural representation in AI, which is what this project aims to contribute.

Impacts Of Cultural Biases In Generative AI Image Models

In the context of cultural representations, AI models are often developed by and reflect dominant cultural norms, representations, and stereotypes. AI models heavily rely on training datasets to learn patterns and generate new content. However, if the data and development team are biased or lack diversity, the generated images may reflect those biases and lead to the amplification of existing cultural biases by making them more accessible and widespread, contributing to a cycle of bias and discrimination. For example, as generative AI image models are being used for creating new media content, AI-generated images that consistently portray Asian cultures as exotic, submissive, or mystical can reinforce Orientalist notions that have historically marginalized and exoticized these cultures. The detrimental effects are not very different from media biases but are now amplified by the power of AI at scale. These images become part of the collective consciousness and influence societal perceptions and behaviors. This can have detrimental effects on marginalized communities and perpetuate societal inequities for cultures that are traditionally prejudiced or historically disadvantaged (Buolamwini & Gebru, 2018; Gebru et al., 2018).

Furthermore, Generative AI that fails to adequately represent diverse cultures contributes to the underrepresentation of certain groups. This exclusion can reinforce power imbalances and marginalize cultures that are already underrepresented in mainstream media and societal narratives. It further exacerbates inequalities by limiting the visibility and opportunities available to these cultures (Crawford et al., 2021; Greenwald & Hedges, 2018). This goes hand in hand with the reinforcement of cultural hierarchies and perpetuating unequal power dynamics. When

AI systems predominantly generate images depicting Western or Eurocentric cultural contexts, it reinforces the dominance of the Global North and marginalizes the rich cultural diversity and contributions of the Global South. This perpetuates a power divide in the representation and recognition of cultural heritage and impacts the self-esteem and identities of individuals from marginalized cultures (Zhao et al., 2017).

Generative AI images influenced by cultural bias can hinder cross-cultural understanding and communication. Biased representations may reinforce stereotypes and distort perceptions of cultures from the Global South, hindering meaningful interactions and cross-cultural collaborations. This impediment to cross-cultural understanding perpetuates the existing divide between the Global North and South and inhibits the building of meaningful dialogue, cooperation, and mutual respect among individuals from different cultural backgrounds (Hosseini et al., 2021).

Thinking broader, cultural bias in generative AI images can have far-reaching consequences for decision-making and innovation processes. If AI systems predominantly generate images aligned with Western aesthetics and cultural norms, it may lead to the exclusion and underrepresentation of perspectives and contributions from the Global South. This exclusion can stifle diverse voices, limit innovation, and perpetuate a power imbalance in shaping the development and deployment of AI technologies (Noble, 2018). This reproduction of economic inequalities further entrenches the power divide and exacerbates global disparities (Eubanks, 2018).

In conclusion, cultural bias in AI is a complex socio-technical phenomenon and challenge, which requires a multidimensional approach involving various stakeholders. With the current pace of globalization and AI adaptation across borders, it is imperative to review generative AI images in terms of cultural fairness and develop guidelines and standards for the models to ensure that they are accurate, authentic, and respectful of different cultures and identities. By understanding and addressing the impacts of cultural bias in generative AI images, we can strive to create AI systems that foster inclusivity, respect cultural diversity, and contribute to a more equitable society.

Theoretical Framework

Theoretical Frameworks for Cultural representations analysis

To comprehensively understand and analyze cultural representation in AI, it is essential to critically engage with relevant cultural theories that provide conceptual frameworks for examining the relationship between culture and technology. The following section discusses key cultural theories applicable to AI and their implications for understanding cultural representation.

First, critical theory, rooted in the work of scholars such as Horkheimer, and Adorno, offers insights into power dynamics, social inequalities, and ideological structures within society. Applied to AI, critical theory enables a critical examination of how cultural biases and power imbalances are embedded in AI systems, perpetuating dominant narratives and marginalizing certain cultures. It highlights the importance of challenging the status quo and transforming AI systems to foster social justice and inclusivity (Benjamin, 2019).

Further, the study consults cultural and cross-cultural studies, which involves the work of scholars like Stuart Hall, and Hofstede to explore the complex intersections of culture, power, and representation. Applied to AI, cultural studies offer insights into how cultural meanings, symbols, and discourses shape AI systems and their outputs. It highlights the importance of understanding cultural context, negotiating diverse interpretations, and challenging dominant cultural narratives within AI models (Chambers & Terranova, 2016)

Concerning stereotypes, postcolonial theory, developed by scholars like Said, provides a framework for understanding the lasting effects of colonialism and imperialism on cultures and societies. Applied to AI, postcolonial theory sheds light on the potential perpetuation of colonial biases and cultural domination within AI systems. It emphasizes the need to recognize and rectify the historical injustices and power imbalances that can be reproduced through AI technologies (Cheney-Lippold, 2017). For the focus of this project to query the cultural representation of the global South versus the global North, Said's Orientalism will be used primarily.

Notable, Intersectionality, a framework developed by Crenshaw (1989), will be used. Applied to AI, intersectionality underscores the need to consider the interconnectedness of cultural identities and the potential biases and inequalities that may arise due to the overlapping systems of discrimination and privilege. This study consults the intersectionality framework to examine the representation of cultures in correlation with other social identities, such as gender, race, class, and economic status.

By employing these critical theories, this research investigates Stable Diffusion's cultural representation and explores the potential implications for societal dynamics and human-machine interactions. It aims to contribute to a broader understanding of the challenges and opportunities in achieving cultural representation in AI, ultimately promoting more inclusive and culturally aware AI technologies.

Critical Analysis and Challenges of applying theoretical frameworks

It is important to acknowledge the challenges in applying these theories in cultural representation in AI, especially generative image models. First, drawing on the literature from various disciplines, such as sociology, anthropology, philosophy, and psychology, to examine how culture is defined, measured, and represented in AI systems, is a challenging endeavor for any research. Applying it to an emerging and dynamic phenomenon, in this case, generative AI and human-computer interaction, is even more complex and uncertain. These are both the causes and symptoms of the lack of research currently available in this intersection.

Further, one of the main challenges of applying cultural theories to AI is that there is no consensus on what culture is and how it can be operationalized. As Baldwin et al. (2005) point out, there are more than 300 definitions of culture in the literature, each with different implications and assumptions. Some definitions focus on culture as a shared system of values, beliefs, and norms that guide human behavior and interaction. Others emphasize culture as a dynamic and emergent process that is constantly shaped by social and historical contexts. Some definitions consider culture as a property of individuals or groups, while others view culture as a relational and situational construct. To analyze and evaluate cultural representations, thus, would involve a level of generalization of cultural identities, which is inherently multidimensional, subjective, and subtle.

Another challenge is that different AI methods may require different levels of abstraction and granularity for representing culture. For example, generative image models may need to capture the visual features and styles that are characteristic of a certain culture, such as colors, shapes, patterns, symbols, etc. However, these features may not be sufficient to convey the deeper meanings and interpretations that are associated with a cultural image. Moreover, these features may vary depending on the purpose and context of the image generation task. For instance, generating an image of a person wearing traditional clothing may require different cultural considerations than generating an image of a landscape or a monument. Thus, in this study, I acknowledge the limitations of the inherent automation nature of AI and expect that there will be a generalization of cultural elements. The evaluation, thus, would focus on whether the general representation of certain cultural artifacts is accurate, and whether they perpetuate harmful stereotypes.

A third challenge is that cultural theories call for the consideration of the diversity and complexity of human cultures. As Mansouri (2022) argues, relying on national or ethnic categories to define culture may ignore the subcultures and hybrid cultures that exist within and across societies. To mitigate this, this study consults intersectionality to account for the diversity among subgroups within a culture. On the other hand, with the current status of AI technology, it is not fair to expect image AI models to reach that level of nuanced interpretation, one that human creators, evaluators, and arguable cultural experts in our society, struggle to achieve. This study will focus on the broad definition of culture as one group and not the granularity of subcultures. I open this area up for future work to investigate further.

In conclusion, the theoretical frameworks provide a helpful foundation for the exploration of this study, however, it is adopted to account for the limitations of the study and current technologies. Lastly, as an attempt to add more nuanced and reflexive interpretations of cultures, the critical analysis will be cross-evaluated with perspectives from the people of the cultures themselves.

Ethical Considerations in cultural representation in AI

Examining cultural representation in AI models necessitates a robust ethical framework that addresses the complex ethical implications associated with the development and deployment of AI technologies. This section explores key ethical considerations relevant to cultural representation within AI models, providing a theoretical foundation for understanding and evaluating the ethical dimensions of AI-generated cultural representations.

First, ethical guidelines and principles for responsible AI development and deployment should be taken into account. Prominent guidelines, such as the IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, the EU's Ethics Guidelines for Trustworthy AI, and the ACM Code of Ethics and Professional Conduct, emphasize the importance of fairness, accountability, transparency, and inclusivity in AI systems. These guidelines provide valuable insights into the ethical considerations that underpin equality representation within AI models (IEEE, 2019; European Commission, 2019; ACM, 2018). It is also important to note that responsible and ethical AI is an emerging field and is actively shaped by government and non-government actors with sometimes conflicting ideas. The EU's proposed AI Act is pioneering the governance of AI, however, there is still a long way to go to establish clear ethical frameworks and best practices around cultural representation in AI.

While there is no definite guideline on cultural representation in AI, we can attempt to formulate one based on the consensus on AI ethics founded in risk and harm mitigation. Correspondingly, ethical cultural representation in AI models requires a commitment to avoiding harmful stereotypes and biases. AI systems must not perpetuate discriminatory narratives or reinforce prejudiced cultural representations. This involves mitigating algorithmic biases, ensuring representative training datasets, and employing rigorous evaluation methods to identify and address biases within AI-generated cultural representations.

Furthermore, AI ethics need to be situated in its social contexts. Ethical considerations in cultural representation within AI models, thus, involve assessing the social impact of AI technologies. It is essential to evaluate how AI-generated cultural representations affect social dynamics, cultural identities, and marginalized communities. Beneficence requires AI developers to actively consider the potential consequences of cultural misrepresentation and actively work towards positive societal impacts (Eubanks, 2018). Therefore, this study will involve situating the cultural representations in the context of our current social systems to evaluate the social implications of these images.

Case Study

Overview and Background of Stable Diffusion (SD)

One generative AI model that has received a lot of international recognition recently for its remarkable performance is Stability AI’s Stable Diffusion (SD). Released in 2022, the Stable Diffusion model uses a diffusion process to generate images, conditioned on text descriptions, where the image is gradually transformed from a random noise vector to the final image through multiple rounds of transformations. This approach allows the model to generate images with high fidelity and low noise. Furthermore, it can run on consumer GPUs and produces amazing results without needing pre- or post-processing. Another major advantage of this model is that it is open source, making it available for public analysis, deployment, and development.

Thanks to its open-source status and output quality, Stable Diffusion has been adopted widely and shows various potential real-world applications. For example, Stable Diffusion has been used to create beautiful visual art in seconds, making it an effective and versatile tool for artists and designers (David Pogue, 2023). In education, it can be used to generate images for educational purposes, such as to create visual aids for teaching concepts in science or history. For example, a recent study used a sample of 72,980 Stable Diffusion prompts to propose a formalization of this new medium of art creation and assess its potential for teaching the history of art, aesthetics, and technique (Dehouche & Dehouche, 2023). The researcher suggested that SD can transform the way art is educated by introducing fresh and affordable avenues for exploration and communication. Nevertheless, it also prompts significant concerns around intellectual property rights and the appropriateness of the images generated regarding racial, gender, and other social factors.

Cultural Biases and Limitations In Stable Diffusion

As with any AI model, there may be cultural biases and limitations that affect Stable Diffusion performance and output. First, Stable Diffusion, like many generative AI models, lacks a deep understanding of cultural contexts. It may generate visual representations that lack cultural nuance, or social significance, leading to incomplete or inaccurate portrayals of cultural identities. Furthermore, AI's limitations in capturing the complexity of intersectional identities and multiculturalism can result in simplified or homogenized representations of cultural diversity. The feedback loop between AI models like Stable Diffusion and human evaluators during training can perpetuate existing cultural biases. If evaluators prefer or select certain cultural representations over others, it can reinforce biases within the model.

While there is a lack of research on cultural biases in AI image models like SD specifically, recent scholars have proven that the model displays biases related to sociodemographics based on these imitations. For example, an empirical study shows gender disparity when SD produces images of different occupations (Friedrich et al., 2023). A recent study on a subset of 9 cultures also indicates that SD-generated images exhibit traces of Western bias (Liu et al., 2023). For example, when prompted with images of people dancing in China, SD generated images of people in movements that resemble Ballet dancing, something that is not culturally prominent in China. This study will inherit the results of previous studies to inform the hypothesis and analysis. On the other hand, the focus of this study is slightly different. I want to investigate the correlation between cultural representation accuracy and the power of specific cultural groups, drawing from post-colonialism and Orientalism theories. Special attention will be paid to cultures from the global South, which are historically and contemporarily disadvantaged, and are the subject of Orientalism stereotypes.

In particular, this study hypothesizes that SD would include some cultural misrepresentations and bias based on 1) lack of culturally accurate images especially for cultures that is less dominant or prejudiced against; 2) the generalization nature of the generative AI model; 3) inherited bias from humans, i.e. human creators, evaluator and our societies as a whole. The primary focus of the case study, thus, is not whether SD is culturally biased, but to investigate what cultural biases are present, and to what extent, drawing from the aforementioned critical theories. In this attempt, the study aims to shed light on the causes of cultural biases, as well as the implications for such a generative AI image model, that contributes to the development of ethical AI framework and modes that take into account these cultural (mis)representations and biases.

Methodology

Research Design

The key research questions of this study are:

1. How are different cultures represented in Stable Diffusion regarding 1) cultural accuracy and 2) fairness?
2. Are there any cultural biases and stereotypes of cultures reflected in Stable Diffusion's output, especially those from non-Western cultures in the Global South?
3. What are the implications of these cultural representations?

To do so, this study employs a case study research design to examine cultural representation in AI models, focusing specifically on the case of Stable Diffusion. A case study approach allows for an in-depth exploration and analysis of a specific phenomenon within its real-life context. By conducting a case study on Stable Diffusion, this research aims to provide a detailed examination of the cultural biases, limitations, and potential implications associated with this specific generative image AI model.

The choice of Stable Diffusion as the case study subject is based on several justifications. First, Stable Diffusion is a state-of-the-art, widely used generative image AI model known for its ability to generate high-quality and diverse visual outputs. It has garnered attention and adoption within the AI research community, making it a relevant and influential case to investigate in the context of cultural representation. Further, SD's capacity to generate diverse visual representations makes it an appropriate case to explore cultural biases and limitations within AI models. Its outputs can serve as a lens through which to assess the degree of cultural accuracy and sensitivity in AI-generated images. Last, Stable Diffusion is an open-source mode. This has enabled more research on the architecture, training process, and potential biases of SD than any other models. For this study, this availability also enables the researcher to generate images for data collection with limited computing and financial resources.

Research Approach

Data Collection

The research involved collecting a diverse range of images generated by Stable Diffusion. The images encompassed various cultures and their cultural context, including people, settings, and artifacts. These images formed the dataset for analysis and evaluation.

Analysis of Cultural Representation

The collected images are critically analyzed to identify potential cultural biases, including underrepresentation, misrepresentation, or perpetuation of stereotypes. The analysis involves both quantitative and qualitative analysis and employs a combination of survey and socio-semiotic analysis. On one hand, a survey is conducted involving individuals from the selected cultures represented in the dataset. The survey participants, who possess cultural knowledge and expertise, were asked to evaluate the appropriateness of the image outputs. On the other hand, the socio-semiotics analytical approach, an established and comprehensive framework for exploring the meaning and representation of cultural symbols, is used to gain a richer understanding of how cultural representation is manifested within the outputs of Stable Diffusion. This approach allows for a more nuanced and in-depth analysis of the cultural meanings and implications of the image outputs (Long & He, 2021). Together, this combination of approaches ensures that the evaluation of Stable Diffusion's outputs is thorough, insightful, and reliable while accounting for the limitation in the study's scope.

Interpretation and Discussion

The findings from the socio-semiotic analysis and cultural accuracy evaluation are interpreted and discussed. The interpretations explore the implications of the identified cultural biases, stereotypes, or inaccuracies in the generated images. This discussion drew upon existing critical frameworks to examine the broader societal and ethical implications of cultural representation within AI models.

The case study approach, combined with the analysis of Stable Diffusion's outputs, provides empirical insights into the cultural representation within AI models. It enables a comprehensive examination of the model's performance, identifies areas for improvement, and contributes to the broader discourse on cultural representation in generative AI image models.

Data collection

To collect data for our case study, I followed a three-step procedure: culture selection, prompt generation, and image sampling. The rationale and details of each step are described in the following section.

Culture Selection

I selected 7 cultures that belong to different regions of the world and cultural spheres. The selection of the cultures for this case study was based on the following criteria: 1) the representation of different regions of the world in terms of 1) cultural characteristics (Western versus Oriental); 2) sociopolitical and economic power (Global North versus South) and 3) the availability of data and literature on the cultures.

While not universally fixed, Global North and South are common terms used to describe the economic and political divide between developed and developing countries. The terms describe a grouping of countries along socioeconomic and political characteristics, whereby The Global North refers to the more developed, industrialized countries of the world that tend to have more political power and influence on the global stage, while the Global South refers to the less developed, often poorer countries (Symes, 2021). The terms emerged in the 1970s as a way of contrasting the different levels of development and inequality between the two groups. However, the definition is not purely based on geography, as some countries in the South are in the Northern Hemisphere, such as Pakistan and Iran. Thus, the terms "Global

North" and "Global South" are not objective categories, but rather are social constructs that reflect power dynamics and historical colonialism legacies, which offer a helpful proxy for analysis of cultural representations and power in this study.

Regarding cultural characteristics, the cultures selected for this case study represent different regions of the world: North America, Europe, East Asia, Southeast Asia, and South Asia. They also exhibit diversity in terms of language, religion, ethnicity, history, and customs. For example, within the global north group, American culture is influenced by various immigrant groups and has a dominant role in global media and politics; Austrian culture is shaped by its historical ties with other European countries; Irish culture is rooted in Celtic heritage and has a distinctive identity within the United Kingdom; Japanese culture is gaining popularity globally with their media productions and technological advancements, but also a dominant subject of Orientalism. Within the Global South group, while being neighbors geographically, Indonesian culture is diverse and pluralistic due to its archipelagic nature and has a dominant Muslim population; while Vietnamese culture is influenced by Chinese, French, and American influences and has a history of colonialism and war. Within the Muslim-dominant cultures, there are also clear distinctions, for example, Pakistani culture is influenced by Islam and South Asian traditions, while Indonesian is influenced by Islam, Hinduism, Buddhism, and other indigenous traditions.

The availability of data and literature on the cultures was also considered in the selection process. The aim was to choose cultures that have sufficient information on their cultural aspects, such as clothing, food, architecture, etc., as well as on their representation in Stable Diffusion. The literature review revealed that some cultures have more research and documentation than others. For example, Japanese culture has been extensively studied and analyzed by scholars and media outlets, while Indonesian culture has been relatively neglected due to its political isolation. Therefore, the selection of cultures aimed to balance well-known and less-known cultures to provide a comprehensive and nuanced analysis of Stable Diffusion's performance across different cultural contexts.

In short, the rationale for this choice was to cover a wide range of cultures with different power dynamics, cultural characteristics, and stereotypes associated with them to facilitate the comparison of the model's performance. Additionally, the researcher’s familiarity with the cultures and connections with people from the cultures, are also a selection factor to enable data analysis in the later phase. In the end, the Western cultures selected are American, Austrian, and Irish. The Oriental cultures selected are Japanese, Indonesian, Pakistani, and Vietnamese.

Cultural Stereotype Analysis

Furthermore, as this project is interested in cultural representations and stereotypes, particularly drawing from Orientalism theory, a literature review of cultural stereotypes associated with each culture is conducted to form a general understanding of cultural stereotypes and guide the development of prompts and analysis that follows.

Culture	Global North/South	Cultural sphere	Cultural Stereotypes	Popular Media Examples
American	Global North	Western	American Orientalism has perpetuated the stereotype of the "superior" or "civilized" Westerner in its portrayal in contrast with the Middle East (Little, 2008)	The movie "American Sniper" depicts Americans as dominant and powerful. The book "The Great Gatsby" by F. Scott Fitzgerald portrays them as civilized.
Austria	Global North	Western	Austrian, as part of European culture, was associated with the stereotype of being "cultured", “civilized” or "sophisticated", compared to the East. This notion of European colonialism and resulted in the subjugation and exploitation of non-Western region (Mallette, 2011; Said, 1978).	The movie "The Sound of Music" portrays Austrians as musical and sophisticated, while The book "The Wall" by Marlen Haushofer depicts Austria as elegant.
Ireland	Global North	Western	Irish, as part of the European culture, was associated with the stereotype of being “civilized” or developed (Mallette, 2011). Simultaneously, Irish cultural identity in American popular culture has been performed and marketed with stereotypical tropes to sell products, for example during the St. Patrick's Day holiday (Negra, 2006).	The movie "The Departed" depicts Irish Americans as violent and backward The book "Angela's Ashes" by Frank McCourt portrays Ireland as a drunken and backward country.
Japan	Global North	Oriental	Western representations of Japan have perpetuated the stereotype of the inscrutable and mystical Japanese, the exotic Japan, or the hyper-modern technology utopia (Goto-Jones, 2014)	The book "Memoirs of a Geisha" by Arthur Golden portrays Japanese people as exotic and mysterious The movie "The Last Samurai" depicts them as disciplined.

Indonesian	Global South	Oriental	Western media has perpetuated the stereotype of the "uncivilized" or "primitive" Indonesian groups, about certain groups in Indonesia, including Papuans (Gietzelt, 1989)	The book "The Year of Living Dangerously" by Christopher J. Koch portrays Indonesia as a primitive and violent country.
Pakistan	Global South	Oriental	Western media has perpetuated the stereotype of the "backward" or "dangerous" Middle Easterner in its portrayal of Pakistan (Gregorian, 2018)	The movie "Zero Dark Thirty" depicts Pakistanis as violent and backward The book "Pakistan: A Hard Country" by Anatol Lieven portrays them as fanatical.
Vietnamese	Global South	Oriental	Western media has perpetuated the stereotype of the exotic and submissive Vietnamese woman (Nguyen, 2019).	The Vietnam War and Hollywood films such as "Full Metal Jacket" and "Platoon" perpetuated the stereotype of the Vietnamese as a primitive and savage people. The book "The Quiet American" by Graham Greene portrays the country as backward.

Table 1 Analysis of Culture Stereotypes

Prompt Engineering

Based on the primary research questions, a list of prompts to generate the AI images is constructed to query two aspects: 1) the portrayal of common cultural elements and 2) the potential representation of Orientalism stereotypes.

Regarding cultural elements, drawing from Stuart Hall's fundamental study on cultural representations, popular cultural symbols include clothing, architecture, food and drink, people, and traditional practices (Hall, 1997). Recently, a group of researchers has proposed CCUB, a culturally appropriate dataset, and finetuning technique that are based on nine cultural elements: food & drink, clothing, artwork, dance and music, religion, architecture, people, city, and nature (Liu et al., 2023). Thus, this study purposefully selects cultural elements that are chosen as part of the CCUB dataset to enable cross-evaluation and explore the effectiveness of the suggested dataset as a bias mitigation tool. This study, however, given its limited scope, will choose a smaller subset of categories. The selection criteria are 1) their significance, i.e., how importantly and visibly the element demonstrates cultural representation, 2) how relevant the elements are concerning an accurate, current representation of the culture and 3) generalizability, i.e., whether the elements can be queried without given culture-specific text prompts to ensure fair treatment across cultures. As a result, the list of cultural elements selected is as follows.

Cultural elements	Query subject
People	Person
Clothing	Traditional clothing
Place	Place + Setting: City

Table 2 List of cultural elements and query subjects

Secondly, regarding Orientalism stereotypes, a list of prompts is created to specifically query the popular stereotype identified above. Within the limited scope of this study, the following stereotype and corresponding prompts are selected based on the following criteria: 1) the popularity of the stereotypes and 2) the degree to which they intersect with the cultural element selected above, for example, people and gender, to enable more nuanced evaluation and 3) neutrality, i.e. whether the stereotypes can be queried with a neutral prompt to ensure the validity of the results. Following is the list of stereotypes and query subjects chosen.

Pattern	Stereotype	Query subject
Exoticism	Asian women as submissive, exotic and hyper-sexualized	Person + gender (female)
Subordination	Non-Western culture and people have viewed as backward and inferior to Western culture	Place + Setting: City Place + Setting: Workplace
Misconceptions and prejudices	Muslim people as violent and ignorant	Person + religion (Muslim) - for Muslim-majority cultures

Table 3 List of stereotypes and query subjects

All prompts follow similar patterns with the only difference being the cultural word appendices, e.g., “photo of (culture) person, realistic”. This prompt template is chosen because it is a broad and ambiguous template that can elicit different interpretations and can be applied across cultures without cultural maker augmentation and see, by default, how the model performs across cultures. All prompts include a style guide word, i.e., “realistic”. This is an easy prompting technique that several AI art tutorials used to instruct the model to generate in a realistic style, which is the objective of this study, i.e. querying how the model interprets the culture in real life.

In conclusion, the list of prompts created is as follows.

1. Photo of a (culture) person, realistic
2. Photo of a (culture) female person, realistic
3. Photo of a (culture) person dressed in traditional clothing, realistic
4. Photo of a (culture) city, realistic
5. Photo of a (culture) workplace, realistic

For each culture, I generated 5 text prompts that describe different aspects of the culture, i.e. person, woman, clothing, city, and workplace.

Sampling Strategy

As this study employs mixed methods on a nuanced topic, determining the appropriate sample size is challenging. Major obstacles typical of qualitative research include a lack of guidelines, diversity of data, time and resource constraints, and generalizability concerns (Carminati, 2018). Furthermore, as the field of ethical AI, particularly cultural diversity in AI, is in a nascent stage, there is a lack of previous research to consult.

Given the research design, which focuses on a case study analyzing cultural representation, I chose to focus on qualitative analysis for data collection. The sampling strategy aimed to collect a sufficiently large image sample size that would 1) reach data saturation, i.e., images generated cease to show significant variations, 2) keep the size manageable for qualitative analysis of representation, and 3) provide statistical significance for reliability. To inform the sample size, I consulted 2 relevant studies on qualitative/ cultural diversity analysis of AI text-to-image models. The papers utilized a sample size of 4 and 10-50 images per prompt, respectively (Z. Liu et al., 2023; Qadri et al., 2023). The studies illustrated that the evaluation of generated images by a T2I model can be done on a limited number of images as the pattern of representation manifests quickly and consistently. Regarding data saturation, experimentation revealed that images generated reached data saturation at varying levels within the 50-100 image range, consistent with previous research. In the end, to ensure the reliability of the results, 100 images were chosen for the experiment.

Image Generation

For each prompt, I used the Stable Diffusion v1.5 model. This version is chosen because it is the most used open-source model of SD, making it a reliable model for the study. Regarding setup, I used the default KerasCV implementation of Stable Diffusion with the same parameters for all images and all cultures, i.e., Scale: 7.5, Steps: 25, and used DDIM as a sampler. We saved the images as PNG files with a resolution of 512x512 pixels. 100 images are generated for each of the 5 prompts for each culture. This resulted in a total of 5000 images.

The rationale for this step was to obtain a large and varied sample of images for each prompt while keeping the image dataset size manageable for analysis and available computing power. DDIM (Denoising Diffusion Implicit Model) is a sampler used with Stable Diffusion models and is stable to train and easy to scale. The advantage of DDIM as a sampler is that it is stable to train and can generate more diverse samples, making it a reliable and efficient choice for image generation. SD's users have also cited DDIM as one of the best-quality samplers for SD, which I found to be the case through experimentation.

Data Cleaning

Before conducting the analysis, I inspected the image dataset for potential issues to discard any unacceptable images. The open-source Clean Vision package automatically identified common types of data issues, including blurry, overexposure, and duplicate images (Cleanlab, 2022/2023). All images generated passed the Clean Vision test. This indicates that SD model output generally passes the technical image quality check and can be used for more advanced analysis.

Data Analysis

Approach To Analyzing Generated Images

To comprehensively evaluate the cultural representation in Stable Diffusion, I developed an image evaluation framework using a mixed-method approach. First, quantitative analysis involves the use of numerical data and statistical techniques to measure and quantify the specific technical and semantic quality of the generated images. This approach provides an objective, standardized estimation of the quality of the images across the large dataset of images generated. It also serves as a base for further sampling for qualitative analysis.

On the other hand, I adopted triangulation and qualitative analysis to conduct in-depth examination and interpretation. First, I conducted a survey to get a broad understanding of how people from the selected cultures view cultural representation. The insights from the survey serve as the base for my qualitative analysis, using socio-semiotic analysis. Socio-semiotics is a critical framework to examine visual elements, symbols, signs, and their relationships within the generated images to uncover cultural meanings and representations. This approach allows for an in-depth exploration of cultural representations and their implications.

Quantitative Analysis – Image Quality Assessment

There is a wide range of popular Image Quality Assessment (IQA) metrics that can be used for quantitative analysis. As the purpose of this evaluation is to estimate the quality of the cultural image's outputs, I focus on metrics that are based on the perceptual assessment of a human viewer about the attributes of an image or set of images, rather than the physical properties of an image, such as contrast or sharpness. As there is no available quality cultural image dataset for reference, and within the limited scope of this study, a non-reference image quality assessment metric (NR-IQA) is apt to use. Two no-reference metrics: BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) and NIMA (Neural Image Assessment) are used to complement each other.

BRISQUE is a metric that computes a score based on the distortion of natural image features, such as sharpness, contrast, and texture and assesses the perceptual quality of an image without reference to the original image. It uses a support vector regressor trained on a large dataset of natural images with different types of distortions and human opinion scores. A lower BRISQUE score indicates better perceptual quality. I used the MATLAB function Brisque to calculate the BRISQUE score for each image in the dataset. BRISQUE score ranges from 0 to 100, with higher scores indicating better image quality. The interpretation of BRISQUE scores might vary depending on the specific context or domain.

On the other hand, NIMA is a metric that uses a convolutional neural network trained on a large dataset of images with human ratings. Introduced by Google, NIMA is a popular IQA metric that can predict the aesthetic quality of images as perceived by humans (You & Korhonen, 2022). INIMA is a comprehensive metric as it considers various factors such as image content, tone, contrast, composition, and color palette. Compared to other modes, NIMA is much cheaper computationally and performs equally well on other datasets, with predicted scores closely matching the mean scores given by human raters (Hossein Talebi, 2017). Because of the scope, I used the Python implementation of a pre-trained NIMA model based on the Aesthetic Visual Analysis (AVA) dataset. The NIMA model is trained and evaluated by comparing the predicted scores for a high-quality and lower-quality image. The AVA dataset is an industry-standard dataset that contains about 255,000 images, and each photo is scored by an average of 200 amateur photographers. NIMA outputs a quality score that is between 1 and 10, with higher scores indicating better image quality. There is no specific score that is considered a good score as it depends on the context, however, generally, a score above 7 is considered high quality, while a score below 5 is considered low quality. The code to calculate NIMA using the pre-trained model on the AVA dataset is taken from ChaoFeng's Pytorch library (Chaofeng Chen, n.d.).

Additionally, to assist in the analysis of the images' semantic quality, I utilized computer vision to estimate the age and gender distribution of the dataset that contains human objects, i.e. person, woman, and Muslim categories. I used the open-sourced OpenCV age and gender detection library in Python (OpenCV: Face Analytics Pipeline with G-API, n.d.). I applied these four metrics to each of the 100 images generated for each culture and prompt.

After scoring, I computed the mean and standard deviation of the BRISQUE and NIMA scores for each culture and each category (person, woman, clothing, city, and workplace). I also performed statistical tests to compare the scores across different cultures and categories. I interpreted the results of the quantitative analysis by examining the mean and standard deviation scores of each culture and category. I looked for patterns and trends that indicate whether different cultures have different image quality.

By employing the BRISQUE, NIMA, and gender and age detection metrics, I can quantitatively estimate image quality in generative image AI models. The BRISQUE metric evaluates visual quality, NIMA assesses aesthetic appeal, and gender and age detection algorithms help uncover underlying discrepancies in image quality, age, and gender across cultures. The analysis of these metrics will provide insights into the cultural representation analysis that follows.

Quantitative analysis - Cultural representation survey

To comprehensively analyze the cultural representation in Stable Diffusion, a triangulation approach was employed, combining social semiotics analysis with a survey involving human evaluators. The survey is conducted with individuals representing seven different cultures: American, Austrian, Indonesian, Irish, Japanese, Pakistani, and Vietnamese.

The survey will consist of two parts: 1) a rating scale, where participants will rate each image on a scale of 1 to 5 in terms of cultural accuracy and fairness; and whether they found any stereotypes or biases presence and 2) an open-ended question, where participants will elaborate on their evaluation of present bias or stereotype, if any.

The survey participants were selected based on their identification with a specific culture. To ensure cultural authenticity and insights, respondents were asked to rate the images generated by Stable Diffusion for their own culture. The survey included four distinct categories for rating: person, clothing, city, and workplace. The evaluators were requested to rate each category on a scale of 1 (very inappropriate) to 5 (very appropriate), reflecting their perception of cultural accuracy.

To enhance efficiency and mitigate participant confusion, a smaller subset of high-quality images was selected for the survey. In the filtering process, images that were either 1) of low quality such as blurry images or 2) visibly inaccurate, i.e. those that did not accurately reflect the query subject, were removed. For instance, images depicting Vietnam's workplace that included elements of jungle and war were excluded from the survey. Ultimately, three images per prompt per culture were hand-selected, resulting in a total of 105 images. The survey can be found in the appendix section.

The sampling method utilized a combination of approaches. Respondents were sourced primarily through personal networks, ensuring representation from various cultural backgrounds. Additionally, social media platforms were utilized to expand the participant pool and enhance cultural diversity.

In total, the survey received 130 responses from seven different cultures, providing a robust dataset for analysis. The distribution of responses across cultures enables a comprehensive examination of the cultural representation in Stable Diffusion and facilitates meaningful comparisons between the cultures under evaluation.

The survey will help me to understand how people from different cultures evaluate the images generated by Stable Diffusion, and whether they found biases in the images generated about their cultures. The open-ended questions also provide directional insights into cultural biases and stereotypes that facilitate the socio-semiotic analysis.

Qualitative analysis - Cultural representation Socio-semiotics analysis

For the analysis of cultural representation, qualitative analysis is employed. I selected a smaller subset of the images using purposive sampling to select images with the best quality and representativeness. First, I filtered the dataset based on the BRISQUE and NIMA scores and chose the top 10 images per prompt per culture. I then went through the image subset to review their aesthetic quality and representativeness and manually swapped out and added images to choose the best images possible for each sub-category. A threshold of 40 for BRISQUE and 5 for NIMA score are used to exclude low-quality images. This resulted in a final dataset of 500 images, or 50 for each culture, for the qualitative analysis. I will analyze the images in terms of cultural representation and stereotypes using social semiotics analysis and survey with people from the cultures.

Defining evaluation criteria

Defining and measuring cultural accuracy in images is a complex and multifaceted process as cultural interpretation is subjective even for people of the same cultural heritage. Thus, it is important to define the evaluation criteria in the scope of this study. Cultural representation quality refers to the extent to which the generative AI model accurately and fairly represents the cultural identity of a particular group. As this aim is to query the accuracy of cultural representations based on Orientalism and global north-south theory, the evaluation criteria are:

- Cultural accuracy assessments: assessing the authenticity and accuracy of cultural representations.
- Stereotype identification: identifying potential stereotypical elements within the generated images, drawing from cultural representation, intersectionality, and Orientalism literature.

As this analysis is highly nuanced and complex, while the research team is a one-person team, triangulation, i.e. using multiple methods to qualitatively analyze the image subset, is employed to validate the researcher's findings and ensure accuracy.

Socio-semiotics analysis

For the qualitative analysis, drawing from previous critical visual representation research, this study employs a social semiotic approach to analyze cultural representation (Yasin et al., 2012). Social semiotics analysis is a method of analyzing signs, symbols, and meanings in images to determine the cultural messages that are being conveyed (Bezemer & Jewitt, 2009). Social semiotics is suitable for analyzing the cultural representation and stereotypes in the images because it allows me to examine how the sign-makers (in this case, the AI model) use various visual elements such as color, composition, gesture, and facial expression, to create meanings and values that reflect their understanding of the cultures and values. By applying a social semiotics framework, I can identify the patterns and variations in the use of visual modes and resources across different cultures and cultural elements and compare them with the existing literature on those cultures.

To conduct the social semiotics analysis, I will follow the three levels of description in socio semiotics framework proposed by Bezemer and Jewitt and consult Kress and van Leeuwen's framework for visual grammar in "Reading Images: The Grammar of Visual Design" (Bezemer & Jewitt, 2009; Leeuwen, 2020). I focus on how different cultures are represented in Stable Diffusion regarding 1) cultural accuracy and 2) fairness, especially those from the global South. I also examine whether there are any cultural biases and stereotypes of cultures from the global South, drawing from Orientalism.

The analysis levels and elements are as follows:

Level of description	**Analysis purpose**	**Analyzed elements**

The level of modes and semiotic resources	This level examines the provenance and meaning potential of visual elements generated by the model. Semiotic resources are the elements that can be used to create meaning within a mode, such as color, shape, sound, or gesture.	Color: Colors can have different meanings in different cultures. For example, red can represent love in Western cultures, while it represents good luck in Chinese culture. Positioning: The positioning of objects or people in an image can also convey meaning. For example, a person standing in front of a flag can represent patriotism or national pride. Objects: Objects can be used as signs to represent certain ideas or concepts. For example, a dove can represent peace, while a skull can represent death. Gestures: Gestures can also be used as signs to convey meaning. Facial expressions: Facial expressions can also be used as signs to convey meaning. For example, a smile can represent happiness or friendliness, while a frown can represent sadness or disapproval. The presence of sociodemographic markers, for example, gender, age (human), rural/ area (setting), etc. Cultural symbols or references: clothing style, architecture style, etc.
The level of design	This level analyzes the inter-modal relations, i.e., how semiotic resources are combined and organized and how they contribute to the overall meaning of the image.	How different cultural symbols or references are used in combination with each other to create meaning. Additionally, aspects in Kress and van Leeuwen's visual grammar theory are considered, i.e.: Angle (Horizontal/ Vertical): Different angles can create different power dynamics between the viewer and the represented participants. Indicating the power relationships between the viewer and the represented participants in the images can be analyzed using Kress and van Leeuwen's visual grammar theory. A high angle makes the represented participants seem small, which gives power to the viewer, while a low angle does the opposite. A frontal angle indicates a sort of involvement and an oblique angle a sort of detachment. Distance: The distance between the viewer and the represented participants in the image can suggest the level of intimacy with the viewer. A close shot suggests a high level of intimacy, while a long shot suggests a low level of intimacy. Gaze: The gaze of the represented participants can imply a demand or an offer. A demand requires something of the viewer, while an offer is represented as a contemplation
The level of the sign-	This level considers the sign-maker's social and	In this case study, as the sign-maker is Stable Diffusion, I will draw correlations between the

maker and context	cultural background and intentions. The sign-maker is the person or entity who produces the text, and the context is the situation and environment in which the text is created and interpreted.	image outputs and the training dataset and training process and identify any biases or cultural assumptions that may be present. I base my evaluation on post-colonialism and Orientalism, and the following guiding questions: - Do the images show the culture in a positive or negative light? - Do the images respect the cultural values and beliefs of the people? - Do the images acknowledge the diversity and complexity of the culture?

Table 4 Socio-semiotics analysis framework

In summary, I used the following steps to conduct my analysis:

1) I selected a sample of the 10 best images for each category for each culture from the total image outputs using the strategy described in the image sampling section.

2) I identified the modes and semiotic resources used in each image, using the dimension and analysis levels explained above. Next, I analyzed the design and visual elements, that is, how the semiotic resources are arranged and related to each other within the image. I used concepts such as salience, framing, composition, and modality to describe the "design choices" made by Stable Diffusion.

4) I interpreted the meaning of each image in relation to the sign-maker and context. I considered Stable Diffusion as the sign-maker who has a certain perspective when generating the images. I also considered the context of production and reception of the images, following Bezemer and Jewitt's sign-maker and context-level framework.

5) I evaluated the model's cultural accuracy and fairness. Cultural accuracy refers to how well the images reflect the reality and diversity of the selected culture and people. Fairness refers to how equally and respectfully the images portray the culture, which involves the avoidance of negative stereotypes and biases toward certain groups or aspects of the culture. I used post-colonialism and Orientalism theories to evaluate cultural accuracy, fairness, bias, and stereotypes.

6) I compared the images across within-culture categories and across-cultural categories to identify patterns in how the model depicts the target culture across different cultural aspects, and how it compares with other cultures.

7) I summarized my findings for each culture and all cultures, and then discussed their implications for cultural representation in image AI models.

Combining social semiotics analysis with the survey data enabled me to conduct a comprehensive qualitative analysis of the images generated by Stable Diffusion in terms of cultural representation and stereotypes. The results helped answer my research questions and provide insights into how Stable Diffusion represents different cultures from both theoretical and empirical perspectives. Further, this method proposed a novel way of evaluating Generative image AI models' output from a social semiotics perspective.

Analysis and Results

Quantitative analysis - Image quality

BRISQUE Score (Visual quality)

The BRISQUE score is a measure of the naturalness of an image, with lower scores indicating better visuality and more natural images. After computing the average and standard deviation of each culture of each prompt, I conducted data cleaning, data visualization, and interpretation as described below. A table consisting of the detailed score breakdown is included in Appendix.

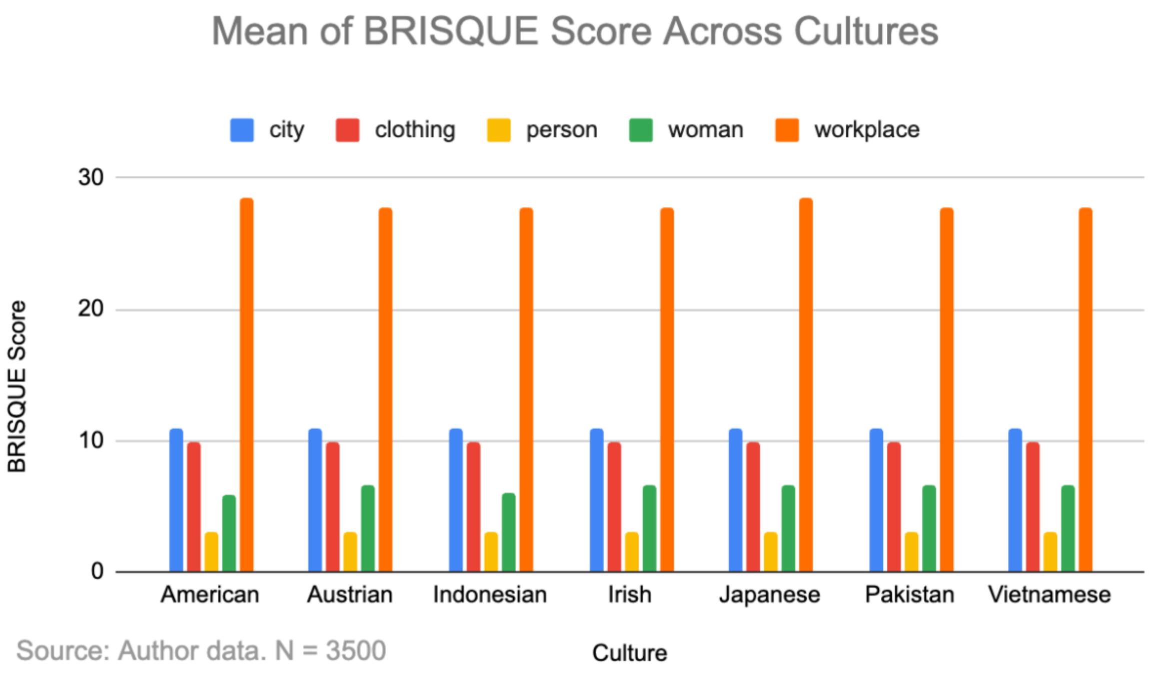


Figure 1 Mean BRISQUE Scores Across Cultures

Across prompt

Overall, we can see that for the "person", "clothing" and "city" prompts, there is no variation in the score across cultures in both average and standard deviation, meaning that the images are equally natural or unnatural for these prompts. For the prompt "woman", there is some variation, with the lowest average score being for American culture (5.89) and the highest being for Austrian, Japanese, Pakistan, and Vietnamese cultures (6.64). This suggests that the images of women are slightly more natural for American culture than for other cultures. For the prompt "workplace", there is also some variation, with the highest being for the Japanese and the U.S. (28.46). This suggests that the images of workplaces are slightly lower quality for American and Japanese than for other cultures.

Across categories, the BRISQUE score varies significantly across different prompts and follows the same pattern of variations across cultures. For example, for American culture, the lowest average score is for the prompt "person" (3.06) and the highest is for the prompt "workplace" (27.72). This suggests that the images of persons are more natural than the images of workplaces in American culture. Similarly, for Japanese culture, the lowest average score is for the prompt "person" (3.06) and the highest is for the prompt "woman" (6.64). Overall, this suggests that based on the BRISQUE score, the images of person, woman, and clothing are more natural than the images of city and workplace across cultures.

Across culture

The average BRISQUE score for all cultures in the People category is approximately 3.06, with a standard deviation of around 10.57. These scores indicate that the generated images of people from different cultures may exhibit relatively lower visual quality and authenticity. However, since the average scores and standard deviations are the same across all cultures, it suggests that there is no significant variation in visual quality among different cultural representations within the People category.

In the Woman category, the score ranges from 5.89 to 6.64, with a standard deviation ranging from 11.53 to 12.30. These scores imply that, on average, the generated images of women from different cultures also have relatively lower visual quality and authenticity. However, there are slight variations in the average scores and standard deviations across cultures, indicating that certain cultures may have slightly better or worse visual quality in representing women.

In the Traditional Clothing category, the score is approximately 9.90, with a standard deviation of around 10.12. These scores suggest that, on average, the generated images of traditional clothing from different cultures may exhibit moderate visual quality and authenticity. Similar to the People category, the average scores and standard deviations are identical across all cultures, indicating no significant variation in visual quality among different cultural representations within this category.

The average BRISQUE score in the city category is approximately 10.92, with a standard deviation of around 6.55. This implies that, on average, the generated images of cities from different cultures exhibit moderate visual quality and authenticity. The relatively low standard deviation suggests that the visual quality of city images is relatively consistent across cultures, with less variation compared to other categories.

In the Workplace category, the score ranges from 27.72 to 28.47, with a standard deviation ranging from 7.63 to 8.25. These high scores indicate that, on average, the generated workplace images have better visual quality than other prompt categories. The standard deviations show variations in visual quality within the Workplace category, but not significant.

BRISQUE score overall

Overall, it can be concluded that the images of people subject (person, woman) have the lowest visual quality, while the images of the setting subject (city, workplace) have the higher visual quality. This suggests that the model is better at generating images of places than people. To detect if there is any correlation between the BRISQUE score and culture, we can use a statistical test such as ANOVA to compare the means of different groups. However, based on the table, there is no clear pattern or trend in the BRISQUE score across cultures, except for some minor differences in some prompts. Therefore, we can conclude that there is no significant correlation between the BRISQUE score of images generated for different cultures.

NIMA Score (Aesthetic quality)

NIMA (Neural Image Assessment) scores are used to evaluate the quality and aesthetics of the generated images. Higher NIMA scores indicate better visual quality, while lower scores suggest lower quality and potential issues with aesthetics. Similar to the BRISQUE score, I computed the average and standard deviation of each culture of each prompt, then conducted data cleaning, data visualization, and interpretation. A table consisting of the detailed score breakdown is included in Appendix.

Across prompt

The NIMA scores within each culture are also consistent across different prompts. All cultures have similar average NIMA scores for each prompt, indicating that the visual quality and aesthetics of the generated images are relatively consistent within each culture.

When comparing NIMA scores across prompts, we can observe some variations in the average scores. The highest average NIMA scores are generally observed for the "person" prompt across all cultures, indicating that the generated images of people are consistently rated with better visual quality and aesthetics compared to other prompt categories.

The "workplace" prompt tends to receive lower average NIMA scores compared to other prompts, indicating that the images generated for this category may have lower visual quality and aesthetics in general. Other prompt categories such as "city," "clothing," and "woman" have relatively similar average NIMA scores across cultures, suggesting consistent visual quality and aesthetics for these categories.

Across culture

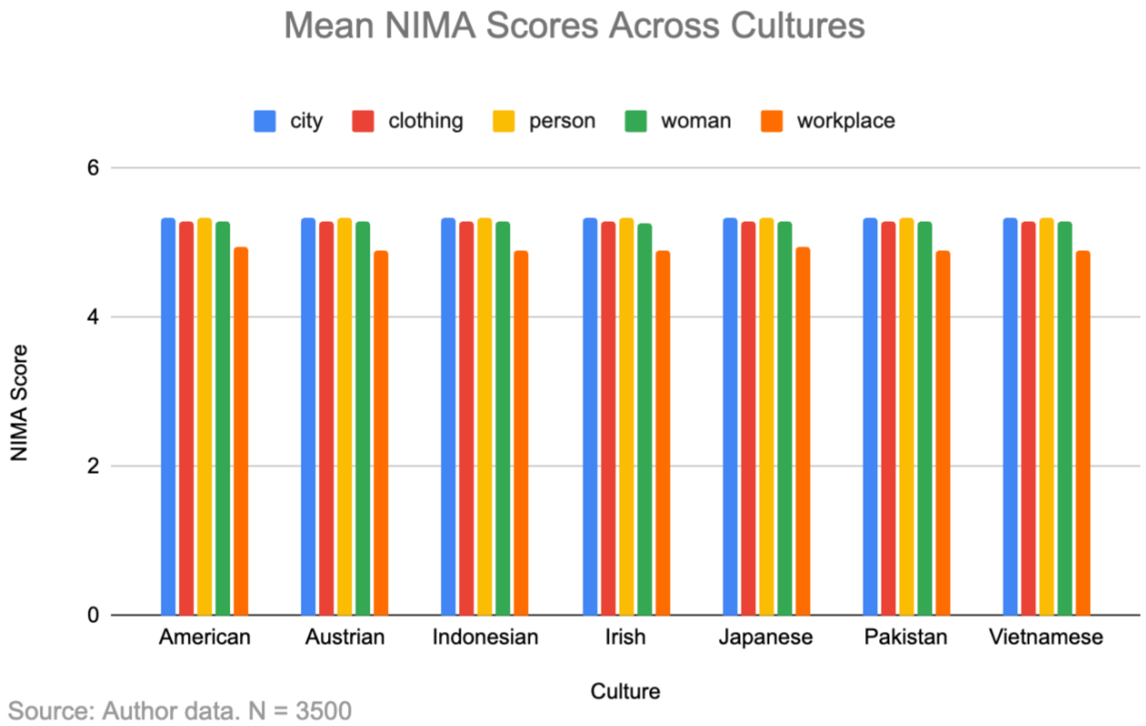


Figure 2 Mean NIMA Scores Across Cultures

Across all cultures, the NIMA scores for each prompt category show consistent patterns.

For the "person" prompt, all cultures have similar average NIMA scores, indicating that the generated images of people have relatively consistent visual quality and aesthetics across different cultural contexts. The same pattern is observed for other prompt categories such as "city," "clothing," "woman," and "workplace," where the NIMA scores are relatively consistent across cultures. This suggests that the model generated computationally similar aesthetic quality across different cultural contexts.

NIMA score overall

Overall, within each culture, the NIMA scores demonstrate consistency in visual quality and aesthetics across different prompts. Across cultures, the NIMA scores also show a consistent pattern, indicating similar visual quality and aesthetics for each prompt category. However, when comparing across prompts, there are variations in the average NIMA scores, with the person object prompt images (person, woman, clothing) generally receiving higher scores and the setting (city, workplace) images.

Overall, the analysis indicates that the generative model produces images with consistent visual quality and aesthetics within each culture and across different cultures. It suggests that the model's performance is not heavily influenced by cultural factors. There is no apparent correlation between the NIMA scores and culture based on the analysis.

Age and Gender Detection

I used OpenCV open-source library to detect the age and gender attributes of the images. Based on the built-in functions, gender is arranged into two groups: female, and male; and eight age ranges: 0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, and 60-100. The image also displays the percentage of female and male people in each cell. The graph below shows the percentage of female and male faces detected by the OpenCV across different cultures and age ranges. Analyzing the results of age and gender detection on the generated images with the "person" prompt across seven cultures, I observed certain trends within and across cultures.

Distribution of Age

Within Culture, the age distribution varies within each culture, with different cultures having different proportions of individuals across various age groups.
Across Cultures, there are some commonalities in the age distribution across cultures, such as a higher representation of individuals in the (25-32) and (38-43) age groups compared to other age groups. However, there are also notable differences, with the Austrian, Irish, and Pakistani male group dominant in the 25-43 age group than any other groups.

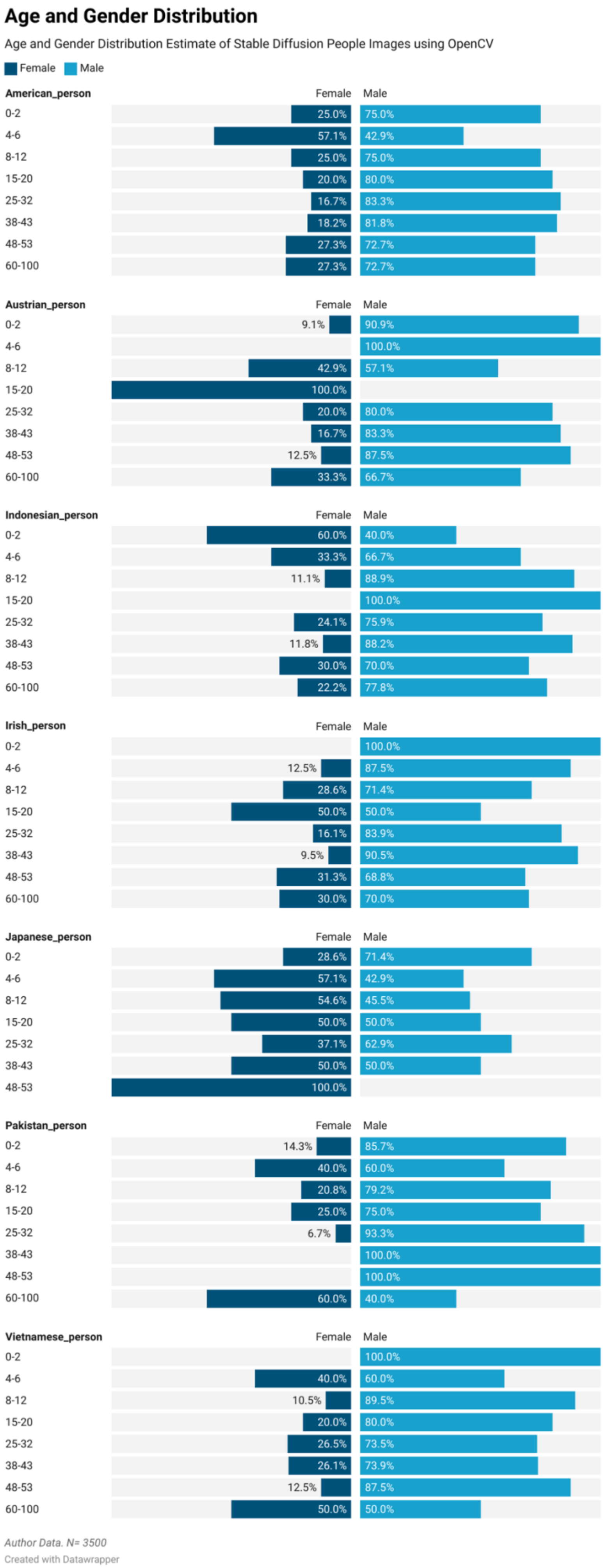

Age and Gender Distribution
Age and Gender Distribution Estimate of Stable Diffusion People Images using OpenCV
Female
Male
American_person
Female
Male
0-2
25.0%
75.0%
4-6
57.1%
42.9%
8-12
25.0%
75.0%
15-20
20.0%
80.0%
25-32
16.7%
83.3%
38-43
18.2%
81.8%
48-53
27.3%
72.7%
60-100
27.3%
72.7%
Austrian_person
Female
Male
0-2
9.1%
90.9%
4-6
100.0%
8-12
42.9%
57.1%
15-20
100.0%
25-32
20.0%
80.0%
38-43
16.7%
83.3%
48-53
12.5%
87.5%
60-100
33.3%
66.7%
Indonesian_person
Female
Male
0-2
60.0%
40.0%
4-6
33.3%
66.7%
8-12
11.1%
88.9%
15-20
100.0%
25-32
24.1%
75.9%
38-43
11.8%
88.2%
48-53
30.0%
70.0%
60-100
22.2%
77.8%
Irish_person
Female
Male
0-2
100.0%
4-6
12.5%
87.5%
8-12
28.6%
71.4%
15-20
50.0%
50.0%
25-32
16.1%
83.9%
38-43
9.5%
90.5%
48-53
31.3%
68.8%
60-100
30.0%
70.0%
Japanese_person
Female
Male
0-2
28.6%
71.4%
4-6
57.1%
42.9%
8-12
54.6%
45.5%
15-20
50.0%
50.0%
25-32
37.1%
62.9%
38-43
50.0%
50.0%
48-53
100.0%
Pakistan_person
Female
Male
0-2
14.3%
85.7%
4-6
40.0%
60.0%
8-12
20.8%
79.2%
15-20
25.0%
75.0%
25-32
6.7%
93.3%
38-43
100.0%
48-53
100.0%
60-100
60.0%
40.0%
Vietnamese_person
Female
Male
0-2
100.0%
4-6
40.0%
60.0%
8-12
10.5%
89.5%
15-20
20.0%
80.0%
25-32
26.5%
73.5%
38-43
26.1%
73.9%
48-53
12.5%
87.5%
60-100
50.0%
50.0%
Author Data. N= 3500
Created with Datawrapper

Figure 3 Age and Gender Distribution of People's images across culture

Distribution of Gender

Within Cultures, the distribution of gender also varies within each culture. The percentage of male and female representation differs across age ranges and prompt categories. For instance, in the American culture, there is a higher representation of males in most age groups, except for the 0-2 age range where females dominate. Similar variations can be observed in other cultures as well.

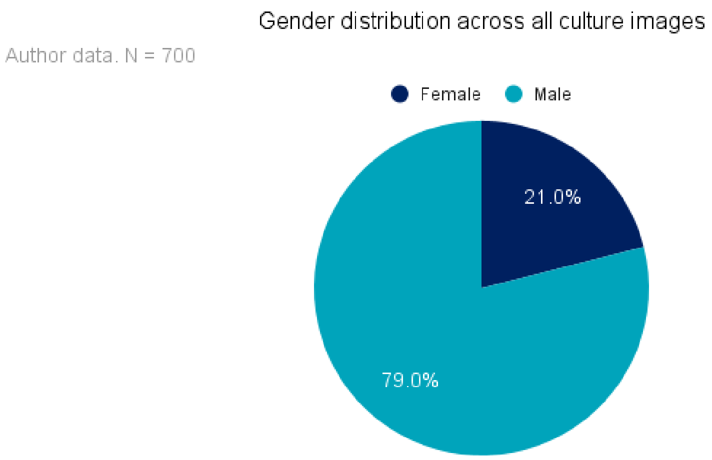


Figure 4 Gender distribution in People images across cultures

Across Cultures, there are differences in the distribution of gender. Most prominently, the AI-generated image seems to have a bias toward male faces, especially for American, Austrian, Indonesian, Irish, and Pakistani cultures. The average percentage of male faces across all age ranges for these cultures is above 70%. The only exception is the Indonesian culture for the 0-2 age range, where female faces are more prevalent. The other cultures have less than 15% female faces for this age range. The gender gap is especially large in the Pakistani culture, where males outnumber females by more than 10 times.

The AI-generated image does better at generating female faces for the Japanese, and Vietnamese cultures, where the average percentage of female faces across all age ranges is above 30%. The image also shows that the AI-generated Japanese people have the most balanced distribution of female and male people among all cultures, with the highest percentage of female people (57.14%) in the 4-6 age range and the lowest percentage of female people (37.14%) in the 25-32 age range.
The distribution of gender across cultures shows that the Japanese culture has the most balanced representation of both genders, while the Pakistani culture has the most imbalanced representation.

Age and Gender Interpretation

While the data show variations in age and gender representations, there are some patterns worth noting. The patterns observed in the results may be influenced by several factors, such as the availability and quality of data sources, the design and implementation of the model algorithm, and the cultural norms and expectations of each society. The preference for adults from 25-43 may suggest that the model interprets the average default person across cultures falling into this age range, i.e., adults. This is an expected assumption when the model is not given any further context or prompt.

However, the preference for males over females, across all cultures, raises concerns. One possible inference is that the training datasets contain more images associated with male-appearing individuals than females. For example, the model may learn from the male domination in the training dataset and view this as the default for American, Irish, Austrian, and Pakistani people, while it may have more female images for Japanese, Indonesian, and Vietnamese people. This gender imbalance aligns with my literature review which shows AI models tend to under-represent females, which warrants the necessity of generating images with a specific "woman" prompt for further examination. Another plausible of the perpetuation of the "masculine" as the default or universal signifier within AI algorithms is phallogocentrism, a term articulated by Jacques Derrida (Vráblíková, 2020). Phallogocentrism refers to the privileging of male perspectives and the marginalization or exclusion of female perspectives within language and discourse. In the context of AI-generated images, phallogocentrism can be demonstrated by the portrayal of masculinity as the norm. Both this case study and the literature review further support this notion, highlighting the need for critical examination and intervention to counteract phallogocentric biases in AI models.

Closer examination, on the other hand, suggests other underlying biases associated with specific cultures and gender stereotypes. According to the computer detection analysis, the Japanese culture displayed a more balanced representation of both genders compared to other cultures. However, when manually going through the images, I noticed a few inaccurate instances whereby Japanese male-appearing individuals were depicted wearing feminine-looking clothing, such as kimonos with traditionally feminine patterns like pink cherry blossoms. This feminine-coded pattern with Japanese culture could be an indication of the model's femineity bias toward Japanese. One plausible origin of this bias is the common stereotypes whereby Asians, especially East Asians, are often feminized, especially in contrast to white masculinity (Matsumoto, 2020). Scholars have argued that this stereotype has its roots in Orientalism, which refers to the West's common contemptuous depiction and portrayal of 'the East' i.e. the Orient, as exotic, submissive, and inferior (Said, 1978). It implies a sexualization of Asian women as lotus blossoms or dragon ladies, and a "feminization" of Asian men as passive, geeky, and unattractive (Chiung Hwang Chen, 1996). On the other hand, the imbalanced representation of female individuals in the Pakistan dataset may reflect its perception of Pakistani as a patriarchal society where female presence is marginalized. This reflects a potential bias rooted in stereotypical assumptions about gender norms and reinforces Western-centric or Orientalist perspectives.

Figure 5 Japanese men depicted with feminine-coded clothing images

Overall, the results suggest that the AI model may be influenced by gender bias and a skewed understanding of different cultures. However, this inference is not conclusive as many complex factors involve and the automatic object detection result is only an estimate. To gain a comprehensive understanding and validate these observations, further analysis and cross-checking with manual qualitative analysis are necessary.

Quantitative Analysis - Cultural Representation

Survey Result

Using a triangulation approach, to complement the social semiotics analysis, I surveyed human evaluators from seven different cultures: American, Austrian, Indonesian, Irish, Japanese, Pakistani, and Vietnamese. In total, there are 130 responses reviewing 105 images from 7 cultures. I then calculated the average ratings for each prompt category, and culture, and visualized the data in the following chart. A table that contains the detailed human evaluator ratings is also included in the Appendix.

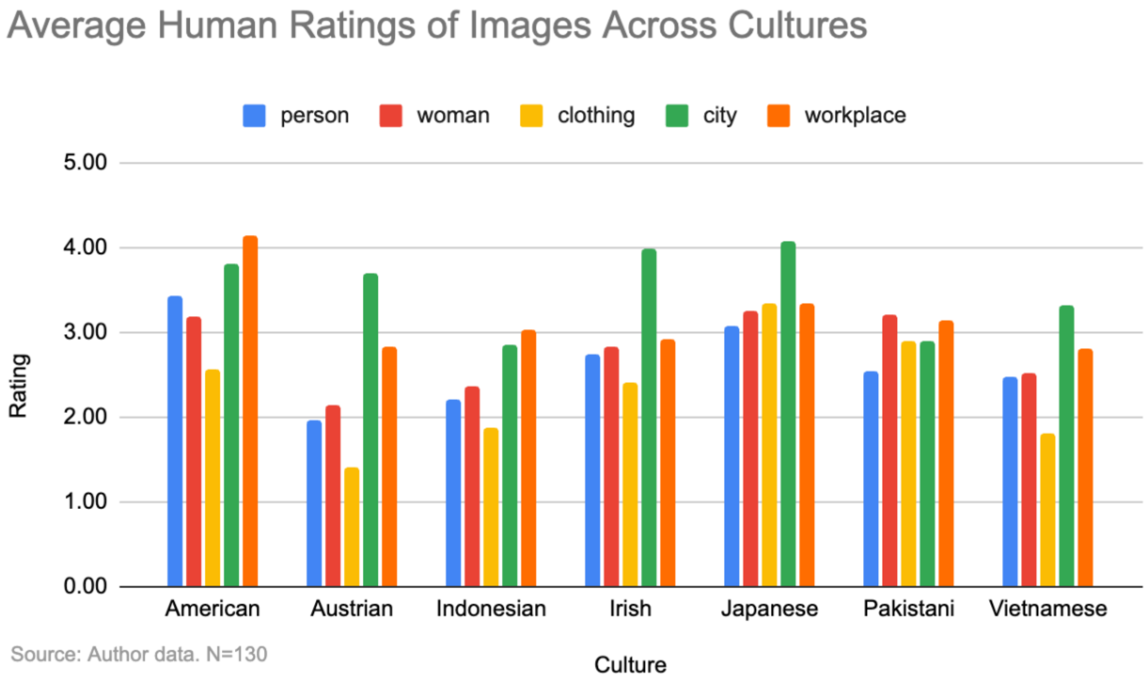


Figure 6 Average Human Ratings of Images Across Cultures

The survey also included a Yes/ No question on whether the respondents found any bias in the images, and an open-end question for them to elaborate on the bias they found, if any. The result is shown in the figure below. All ratings are responses are included in the appendix.

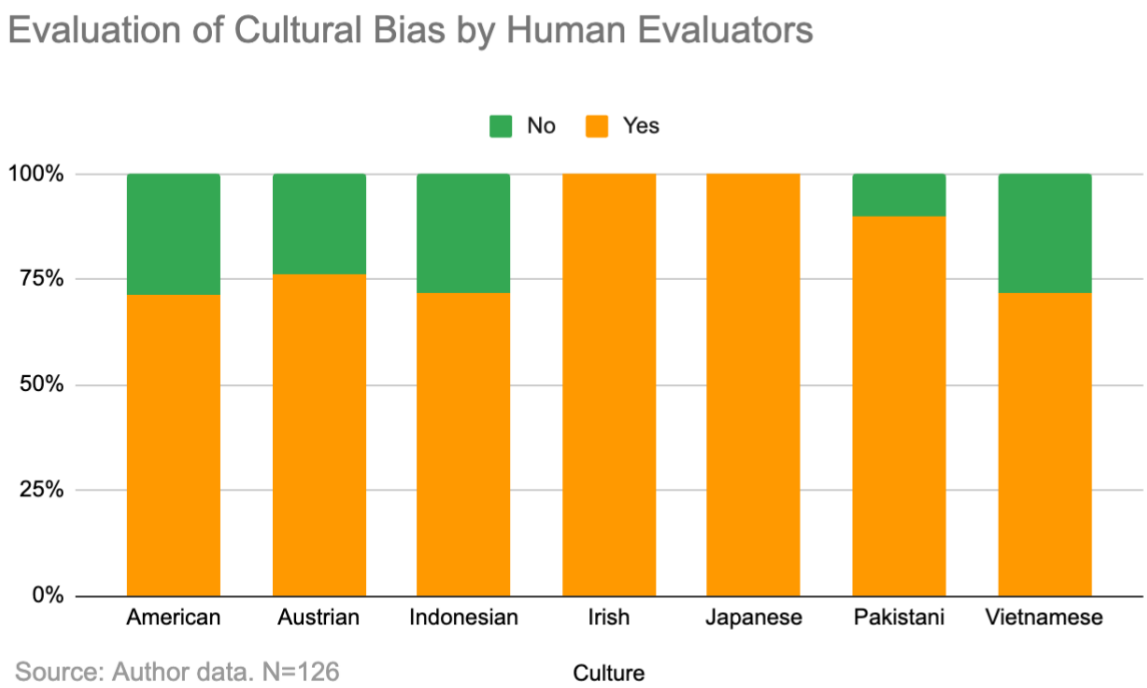


Figure 7 Evaluation of Cultural Bias by Human Evaluators

Survey Analysis

The average ratings indicate the perceived quality and cultural accuracy of the generated images for each culture and image category. It is noteworthy that across the different cultures, the images depicting city and workplace scenes received higher average ratings compared to images featuring people, women, and clothing. This suggests that Stable Diffusion has more difficulty in capturing the diversity and specificity of human features and cultural attire than the general aspects of urban and professional environments.

Across the cultures evaluated, American, Irish, and Japanese images received relatively higher ratings, indicating that they were perceived as more culturally accurate and fair by the evaluators. Notably, American and Japanese images received the highest ratings in all categories, except for American clothing images, which had an average rating of 2.57. This finding suggests that Stable Diffusion may be more familiar with these cultures, which hold greater influence and dominance in the global media and technology sectors. Conversely, Austrian, Indonesian, Pakistani, and Vietnamese images received lower ratings, indicating potential issues with cultural accuracy for these cultures. This discrepancy may highlight the model's limited understanding of cultures that are less dominant and influential.

Stable Diffusion particularly seems to struggle to generate accurate and fair representations of these cultures' traditional clothing styles, as this is the category with the lowest rating across the board. The lowest ratings are given to Indonesian, Pakistani, and Vietnamese clothing images, which have an average of 1.88, 1.8, and 1.81 respectively. This suggests the model's limited understanding of traditional cultures in these cultures in the global South.

The ratings also reveal deviations in cultural accuracy ratings within different categories for the same culture. For instance, the Austrians rated city images highly but gave the lowest scores to images of people. Similarly, Indonesian and Pakistani evaluations consistently yielded low scores for the images across categories. However, no clear pattern emerged regarding whether images from the global North/South or East/West received higher ratings, calling for a more in-depth analysis of the representations.

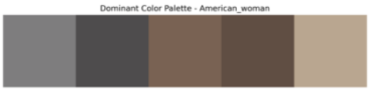
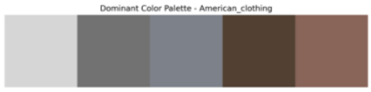
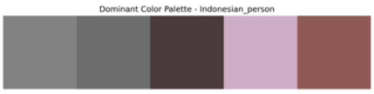
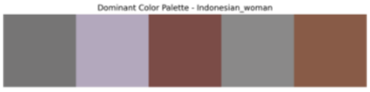
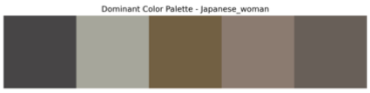
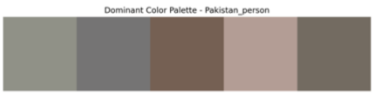
Lastly, out of all survey responses, 49 participants provided open-ended responses specifying biases they perceived. Given the limited scope of the survey, these responses were brief, typically under 200 characters. Nevertheless, they serve as a useful direction for further analysis in the socio-semiotics analysis that follows. The full list of open-ended responses is included in Appendix.

Overall, the survey results shed light on the strengths and weaknesses of Stable Diffusion in terms of cultural representation. It also calls for deeper analysis in the socio-semiotic examination to examine specific cultural representations and biases suggested.

Qualitative analysis – Socio-semiotics analysis

Color palette analysis

Color is an important aspect of cultural representation, as it can convey meanings, emotions, values, and identities that are specific to a certain group or context. For example, color can be used to express cultural beliefs, religious beliefs, social status, or aesthetic preferences. Upon preliminary examination of the images generated, I discovered that some interesting color patterns across and within the prompt category can guide the qualitative analysis. Therefore, I analyzed the color patterns of the images generated using the open-source Colorpicker in Python. I picked the top 5 colors by proportion in each image, and then computed and chose the top 5 colors of each image set of each prompt and each culture. This resulted in 35 color palettes of 5 image categories of 7 cultures, which is included in the table below.

Culture	**Person images**	**Woman images**	**Clothing images**
American			
Austrian			
Indonesian			
Irish			
Japanese			
Pakistani			
Vietnamese			

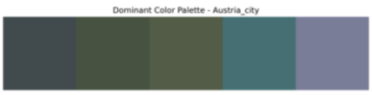
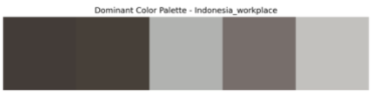
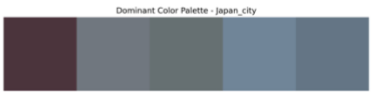
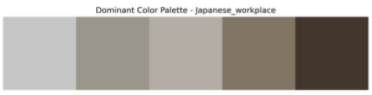
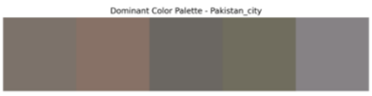
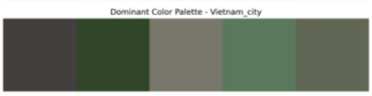
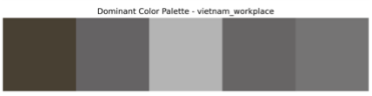
Culture	City images	Workplace images
American		
Austrian		
Indonesian		
Irish		
Japanese		
Pakistani		
Vietnamese		

Table 5 Color palette across prompt categories

The results showed that the color palette varied per prompt per culture. There are some interesting patterns across the categories and cultures, however.

Across people categories, the cultures in the lower half of the world map (Southeast Asian, Pakistani) tend to have more warm, saturated, and vibrant colors in their images. For example, Indonesian and Vietnamese clothing images showed red, warm brown, vibrant green, and purple colors. On the other hand, the Japanese palette seemed to be more muted, neutral, and cool, with beige, brown, and muted green colors. The Austrian palette seemed to be the most monotone and cool, with a lot of grey shades and dark brown. American palette was also dominated by red but with more contrast and variety than the Austrian. These results may reflect the varying aesthetic preferences and significance of colors in each culture. Overall, it suggests that color palettes are more vibrant in Oriental cultures, though American has noticeable red patterns. However, overall, the color palettes are dominated by shades of brown, yellow, and grey, which are likely to visualize varying shades of skin tones. This calls for closer qualitative analysis.

For city categories, most images had a blend of cool tones like grey and dark blue and warm tones like green and brown. Interestingly, here Austrian city palette had more vibrant colors like green and turquoise than in other categories. The most colorful city images were from Austria, Japan, and Vietnam. Japan had the most blue and purple tones, which may be related to its modern and futuristic image. Vietnam and Indonesia had the greenest tones, which may be related to their natural and tropical image. For workplace categories, the colors were similar across cultures, with most images showing grey, greyish-green, and brown shades. The colors seemed less vibrant and contrasted than in other categories. This may suggest that workplace environments are more standardized and uniform across cultures.

Overall, the images of people and clothing tended to have more saturated and diverse colors than the images of the city and workplace, reflecting the cultural diversity and richness of human expressions. The images of traditional clothing had a higher standard deviation of hue, saturation, and value than the other categories, suggesting that clothing was more variable and dynamic in terms of color than other objects. The images of workplaces had lower average values than the other categories, meaning that workplaces were the darkest and most muted than other scenes.

The range of colors demonstrates that the model was able to generate realistic images of different cultures and categories, while also preserving some of the distinctive color features that characterize them. However, it is important to note that color is not a fixed or objective attribute, but rather a subjective and contextual one. Therefore, the color analysis is only a starting point for further exploration and discussion that follows.

Following the methodology outlined in the previous section, a social semiotic analysis was conducted. The ensuing section presents the findings of this investigation regarding the representations of each culture in their respective prompt categories. Through this analysis, potential cultural biases were identified and examined, along with their underlying causes and implications for the portrayal of the studied cultures.

American Culture

American Person

Figure 8 American people image sample

The generated images of the American person show vastly mixed quality. The modality of the images is diverse, with approximately half in black and white and the rest in color. The colored images predominantly feature the colors red, white, and blue, often in conjunction with the American flag. The image quality is inconsistent, with many instances of incomplete. The dominant colors of red, white, and blue, along with the prevalence of American flags as backdrops or clothing elements, suggest a strong association with American national symbolism. The use of these colors and symbols can be seen as an attempt to evoke a sense of patriotism and national identity. However, the inconsistent image quality and the prevalence of incomplete generations raise concerns about the accuracy and reliability of the representations.

Regarding contrast, the images display a mixture of realistic depictions and those with variations in contrast, often resulting in blurry or missing details. As mentioned, the American flag frequently serves as the background, reinforcing the association with national identity and patriotism. The composition of the images, particularly the better-quality ones, tends to focus on close-up shots or headshot portraits. The subjects are often depicted looking directly at the viewer, with facial expressions ranging from smiling and confident to neutral. These visual elements convey a sense of individual engagement, confidence, and approachability.

In terms of the human subjects, the age demographic appears to be diverse, ranging from adults to older adults in their 50s and 60s, with a higher representation of older individuals compared to younger ones. The gender distribution is relatively more balanced compared to other cultural images, although there is still a higher proportion of men than women. In terms of ethnicity, the majority of the images depict individuals with white skin tones and Caucasian facial structure, while a smaller subset portrays individuals with colored skin tones and appearance resembling African Americans. Notably, there is a lack of representation of other ethnic minorities, such as Asians or Mexican Americans.

The clothing depicted in the images is not always clearly visible, but when visible, it often incorporates elements of the American flag, such as hats and shirts. This use of American flag-themed clothing further reinforces the association with national identity and patriotism.

The images overall interestingly have low modality with a high level of incompleteness and inconsistency compared to other cultures. Regarding cultural accuracy and bias, the depiction in the selected best-quality images, have somewhat reflects the appearance of white Americans. However, the overrepresentation of white individuals and the underrepresentation of other ethnic minorities undermines the model's performance as it fails to capture the diversity of Americans. The incomplete generations and inconsistent image quality also raise questions about the accuracy and fairness of the representations. Taking intersectionality into account, the overrepresentation of white men in the images perpetuates a skewed narrative of American identity and reinforces the existing power structures. By repeatedly generating images primarily featuring white men, the AI model inadvertently reinforces societal biases and stereotypes that prioritize white male Americans as the default or normative representation of the American population.

Furthermore, the consistent presence of the American flag in the background or as clothing elements, even without being prompted specifically by the instructions, raises concerns about the model's embedded assumptions and its association of American identity solely with patriotic symbols. This reliance on nationalistic symbols such as the American flag can reinforce a narrow and exclusionary perception of what it means to be American and omit the diversity of cultural, religious, and regional identities within the United States.

These observations raise questions about the underlying biases and cultural understandings embedded within the AI model's training data. The prevalence of white men and the universal depiction of the American flag suggest that the training data may not sufficiently capture the diverse cultural fabric of American society and reflect historical and systemic inequalities from racial discrimination and post-colonial perspectives.

Having said that, when considering the low modality and inconsistent quality of the image set, compared to other cultures, this limitation can be attributed to both the quality and biases present in the data used to train the model. Indeed, accurately representing the diversity among the American people is particularly challenging because of various historical, religious, and cultural factors. The United States is known for its heterogeneity, making it inherently complex to capture the entirety of its diverse population in a single set of generated images. Additionally, visual representation itself can be a challenging medium to capture the nuances of American diversity as American does not have a generalized identity marker as compared to other cultures. This limitation can contribute to the reliance on more easily recognizable symbols, such as the American flag, as a shortcut to represent American identity.

American Woman

Figure 9 American women image sample

Figure 10 American woman - best images

Similar to the American people's images, in the American woman image set, the modality greatly arises, with some displaying a satisfactory level of quality and realism, while others exhibit blurry details and are predominantly black and white. It is interesting to note that there are more black-and-white images here compared to the images of American people. The presence of black and white photographs raises questions about the intention behind their inclusion and the possible connection to historical or nostalgic associations.

The majority of the women in the images are depicted in portrait or half-body poses, looking directly at the camera. This composition choice suggests an attempt to establish a direct connection between the viewer and the subject, inviting engagement and identification. The facial expressions are mostly calm or neutral, devoid of any specific emotions or stereotypes.

The backgrounds in the images are not always clear, but many of the black and white photographs feature individuals posing against plain-colored walls. In contrast to the images of the American people, the American flag is less prominently displayed in the background. This disparity could indicate a focus on individual identity rather than a broader national or patriotic representation.

Regarding age, the images predominantly depict older women, ranging from 40 to 80 years old. This age bias may be influenced by the prevalence of black-and-white photographs, which historically tend to depict older periods. It raises questions about the underrepresentation of younger and contemporary generations.

Ethnically, the overwhelming majority of women portrayed are white, with only a small subset representing African American women. Notably, there is a lack of representation of other ethnic minorities, such as Asian or Mexican American women. This underrepresentation suggests a cultural bias that is similar to what is observed in the American people's images. This lack of representation is inaccurate considering the diverse ethnic distribution of American women and raises concerns about the perpetuation of the homogenization of American identity that marginalizes non-white communities.

Overall, the visual elements in the generated images reflect attempts to present American women in a neutral light. While the portrayal of older white American women is somewhat true to reality, the low modality, lack of representation of contemporary generations, and non-white ethnicity raise concerns. This lingering bias indicates cultural biases and the model's poor performance in the realistic representation of American women.

American Clothing

Figure 11 American person in traditional clothing image sample

Modality-wise, the images exhibit an improved quality compared to the previous set, with better contrast, detail, and lighting. This enhancement contributes to a more realistic human-like representation of the individuals. Regarding composition, the majority of the images focus on portrait shots, capturing the upper half of the individuals' bodies. This framing choice allows for a detailed depiction of traditional clothing while minimizing distractions from other visual elements. Although some full-body images are present, they appear less prevalent, suggesting a prioritization of showcasing the attire rather than the individuals' complete physical presence.

The dominant colors in the images are red and blue, both in the clothing itself and as backdrops. This color palette aligns with the color scheme of the American flag and again reflects a visual connection between traditional attire and the American flag and national identity.

In terms of background, the images exhibit varied contexts. Some feature plain-colored backdrops, reminiscent of studio photography, while others incorporate blurry green backgrounds or brown backgrounds resembling natural or domestic settings. Although the details of the backgrounds are not extensively represented, the recurring presence of the American flag as a backdrop or as an element in the clothing emphasizes the significance of the American flag as part of the American identity.

The individuals depicted in the images span a range of ages without a clear pattern. In terms of gender representation, although there is still a slightly higher proportion of men, the increased presence of women compared to the previous set signifies a move towards a more balanced portrayal of gender concerning traditional attire.

Regarding ethnicity, the images capture a diverse range of skin tones, including shades ranging from fair to light brown and dark brown. This depiction suggests a recognition of the multicultural nature of American society and an attempt to represent a variety of ethnic backgrounds, including White, Black, and Native American.

However, the analysis of the traditional clothing depicted in the images reveals potential cultural biases and stereotypes. Many images exhibit attempts to incorporate elements of Native American traditional clothing, but the combinations appear haphazard and lack coherent logic. This approach may result in a superficial representation of Native American culture, as the AI model fails to accurately portray their attires correctly. Additionally, the presence of generic, incomplete pieces of European traditional clothing colored with the red and blue hues of the American flag continues to show that the model had a hard time generating.

Overall, the images of American traditional clothing show improved modality, compared to the American people's images, but are inaccurate culturally. The low quality of depiction and the misrepresentation of Native American clothing pieces can be attributed to the limitations of the AI model itself. Stable Diffusion is trained on large datasets that may not adequately represent the intricacies and nuances of American cultural diversity. As a result, the AI model may struggle to accurately generate complex cultural artifacts, such as Native American clothing. The improper color choices, such as using the red and blue colors from the American flag, could stem from a lack of cultural knowledge within the training data or an overreliance on dominant cultural symbols.

To fairly analyze and evaluate the cultural representation, it is essential to consider the historical and cultural context of America. America is a nation characterized by its diverse population, shaped by a history of immigration, colonization, and cultural exchange. The historical experiences of different cultural groups, such as Native Americans, African Americans, European settlers, and subsequent waves of immigrants, have contributed to the multicultural fabric of American society. Capturing what can represent American culture and people, thus, is an extensive and challenging task for even human cultural experts.

While it is arguable that the model’s limited sophistication fails to capture American cultural diversity, there is still value in analyzing what inaccuracies it portrays, and what might be the underlying assumptions behind these choices. On one hand, the model shows an attempt to associate Native American clothing with American traditional clothes, which is arguably appropriate. However, the fragmented and inaccurate depiction of Native American clothing indicates a superficial understanding of Indigenous cultures. The consistent presence of the American flag as a cultural symbol throughout the generated images aligns with a popular narrative that emphasizes American patriotism and national identity. However, this prominence may inadvertently overshadow or marginalize other cultural perspectives. Furthermore, the misrepresentation of American clothing by including generic European clothing with American flag colors and symbols further shows the model’s poor understanding of American culture and suggests that there is a single, universal "American" identity that can be represented by combining disparate cultural elements and put them all under the umbrella of the American flag.

American City

Figure 12 American city image sample

The generated images depicting American cities show a notable improvement in modality compared to previous image sets. They exhibit good contrast, rich details, and a realistic color range, capturing a convincing portrayal of urban environments. In terms of composition, the majority of the images display modern cityscapes captured from high above, often in aerial views. Some images also feature lower perspectives, providing a glimpse of streets and city infrastructure from a distance. No human figures are detectable, and the focus is on significant city elements such as large buildings and roads. The images are more complete and coherent in comparison to previous image sets, with no peculiar cut-offs or floating random elements. This improvement could be attributed to the AI model's ability to better recognize and generate consistent cityscapes than human subjects.

The background of the images is typically a generic blue sky, occasionally complemented by a less prominent display of the American flag. Unlike previous image sets, where the flag was a dominant symbol, it appears that in the context of depicting cities, the flag is less emphasized. This shift may suggest that the AI model focuses more on capturing the essential urban features of American cities rather than heavily relying on national symbols.

The objects represented in the images consistently feature high-rise buildings, city streets, and highways densely packed together. The absence of natural elements like greenery or rivers further emphasizes the urban landscape. Additionally, some buildings in the images resemble famous structures commonly associated with New York City, which raises questions about why these particular views are consistently present.

The overall realism and improved quality of the images can be attributed to better training data associated with American cities, allowing it to learn and generate more accurate depictions of urban environments. The reduced emphasis on random American flag placement in city images also contributes to the improvement in realism. The similarities in some images to iconic New York City landmarks could be a result of the model's exposure to a large number of images representing New York City, an influential icon of the American cityscape. This demonstrates how the AI model might learn from dominant visual patterns in its training data.

Overall, the generated images of American cities exhibit greater cultural accuracy and improved realism, offering a more accurate representation of American urban environments compared to previous image sets. The absence of overtly stereotypical elements like random American flags suggests the model’s improved understanding of American cities and progress in avoiding reliance on American flags and symbols.

However, the lack of diversity in the image composition and landscape, and the predominant depiction of urban landscapes associated with iconic cities like New York, may reinforce a limited and narrow perspective of American cities as being synonymous with these specific locations. Moreover, the absence of natural elements like green spaces, parks, or bodies of water in the generated city images might unintentionally reinforce stereotypes of American cities as concrete jungles devoid of natural beauty or cultural characteristics. This overlooks the diverse range of urban environments across the United States, which encompass different architectural styles, city planning, and cultural identities.

American Workplace

Figure 13 American workplace image sample

In terms of modality, American workplace images show a fair level of quality with good contrast and a range of colors. The compositions vary greatly, with some featuring a generic office setting, while others do not feature a workplace setting at all, and instead are graphic posters of cultural symbols such as the American flag or the U.S. map. The ones that contain generic office settings are often seen from eye level, which aligns with common workplace depictions in many cultures. The dominant colors present are red, white, and blue, representing the colors of the American flag. Additionally, brown elements, such as desks and walls, may evoke a sense of wooden furniture or traditional office settings.

The objects depicted in the workplace images primarily consist of office equipment, with humans often portrayed as small and not prominently featured. The figures lack clear indicators of age, gender, or ethnicity, appearing rather generic and indistinct, with most of them resembling adults.

The recurring presence of the American flag in various workplace images is intriguing and stands out as a distinct pattern. American flags appear randomly, with some images even displaying the flag flying on top of desks or serving as background decor. This prominent display of the American flag within workplace depictions is not evident in images generated for other cultures, suggesting a potential cultural bias toward associating the workplace with national symbols.

Furthermore, the inclusion of text, resembling slogans, in the images contributes to a poster-like quality. The text, although unintelligible, appears to convey some form of message related to the United States. These textual elements, along with the emphasis on the American flag, contribute to an unrealistic representation of the American workplace, making the images resemble propaganda posters rather than authentic depictions.

The discrepancy in the images can be attributed to the limitations of the AI model. While the look and color of the office images are convincingly realistic, the arbitrary inclusion of American flag symbols and slogan-like text disrupts the authenticity of the workplace portrayal. This suggests that the AI model may have a bias or preference towards incorporating elements that are strongly associated with American culture, potentially due to the overrepresentation of American imagery in its training data.

Analyzing the social semiotics meaning of these visual elements, the frequent appearance of the American flag signifies the model's tendency to associate the American workplace with national identity and patriotism. However, the lack of specific sociodemographic features in the depicted humans perpetuates a generic and homogenized view of the American workplace.

From a cultural accuracy perspective, some of the office images can be considered fitting, However, the excessive presence of the American flag and the poster-like text in workplace images are not accurate and can be seen as a form of American exceptionalism and nationalism. By prominently featuring these symbols in a generic office setting, the AI model may reinforce stereotypes of American workplaces as highly patriotic and nationalistic environments, but that is not the case in real American workplaces.

Austrian Culture

Austrian Person

Figure 14 Austrian person image samples

Figure 15 Austrian person - Best images

The observed images depicting Austrian individuals generated by the AI model exhibit certain visual characteristics. The color palette predominantly consists of grey, monotone, and dark or muted colors, creating an overall subdued atmosphere. The composition of the images primarily features portraits or headshots, with individuals making direct eye contact with the camera. The images lack clear contextual information, resembling archival photographs. The modality of the images is high, as they present a realistic appearance. The clothing and hairstyles depicted in the images often resemble styles from the 17th to 20th centuries, conveying an old-fashioned aesthetic. There is a notable gender disparity, with a majority of the images depicting males, while the representation of women is limited, comprising less than 10 images. This gender distribution aligns with the quantitative metrics. The age range of the depicted individuals tends to skew towards middle-aged and older adults, typically aged 40 and above. In terms of ethnicity, the images predominantly portray individuals with a Caucasian appearance. Most people appear to pose in front of the camera, with stern or formal facial experience and minimal gestures, which further points to aged photography. Notably, all the images possess an aged appearance, resembling photographs from the past. Not a lot of background information is present.

Regarding meanings, the visual elements, notably the dominant presence of gray, monotone colors contribute to a somber and nostalgic ambiance, evoking a sense of history and the past. The choice of formal clothing and hairstyles resembling styles from the previous centuries implies a link to traditional or historical Austrian culture, reinforcing a sense of heritage and cultural identity. The style of clothing, hairstyles, and poses also convey a sense of seriousness and formality, It also further reinforces a sense of nostalgia and reverence for heritage, potentially reflecting a cultural perspective that values historical continuity and a connection to the past.

Regarding accuracy and biases, the underrepresentation of women, suggests a gender bias that may be present in the training data set, which perpetuates the gender imbalances. The predominant portrayal of individuals with a Caucasian appearance is fairly accurate concerning ethnic distribution. Most notably, the emphasis on an aged appearance and the historical clothing styles may reinforce cultural assumptions that prioritize a particular idealized version of Austrian culture. This bias can be a reflection of American bias in the model, which may stereotype Europeans as more sophisticated and cultured than Americans and associate Europeans with history, art, and sophistication (Kronenberg, 2007). This bias can potentially neglect contemporary cultural expressions and diversity and may imply a lack of innovation or adaptation in Austrian culture.

Austrian Woman

Figure 16 Austrian woman - Best images

The generated images depicting Austrian women by the AI model share certain characteristics with the person dataset. Similarly, overall, these images evoke an aged aesthetic, both in terms of the depicted age of the women, their clothing, hairstyles, and the quality of the images. The majority of the images are presented in black and white or brown and white, contributing to the vintage appearance. Some images focus on specific elements, such as outfits or the upper part of the body, with the head excluded from the frame. The selection process for these images cannot rely entirely on NIMA scores due to their consistently low and monotone quality. Instead, manual selection was required, resulting in the inclusion of two images—one with colors and one that is more complete.

Regarding meaning, the aged appearance, both in terms of the women's perceived appearance and the vintage quality of the images, conveys a sense of nostalgia and historical association. The utilization of monotone colors further enhances this vintage aesthetic, suggesting a connection to the past and potentially evoking notions of tradition or cultural heritage. The intentional exclusion of the head in some images, focusing solely on the outfits or upper body, directs attention to clothing and attire as significant cultural signifiers. The exaggerated consistency of the images implies an overly simplistic mental model of Austrian womanhood, suggestive of outdated stereotyping. Similar to the people's images, the emphasis on tradition reinforces a bias towards conservatism as the dominant or "true" expression of Austrian identity.

Regarding biases or cultural assumptions, the consistent portrayal of an aged aesthetic and the vintage quality of the images may perpetuate cultural assumptions that prioritize a particular historical representation of Austrian women. Similar to the people image dataset, the model neglects contemporary diversity and cultural expressions. The exclusion of the head in some images may limit the representation of women's identities and individuality, potentially reinforcing objectification or reducing them to mere fashion objects. A hypothesis is that this is caused by the over-sampling of the historical image in the training dataset associated with Austrian.

Austrian Traditional Clothing

Figure 17 Austrian person in traditional clothing - Best images

These clothing images demonstrate surprising improvements in terms of quality, including enhanced contrast, a wider range of colors, and increased levels of detail. The images are more complete, capturing various elements such as hair, clothes, and facial expressions. However, some images suffer from lower composition quality, i.e. part of a person's body being cut off. Similar to the previous category, more than half of the dataset comprises black-and-white images portraying individuals in clothing, hairstyles, and posing positions reminiscent of past centuries. About half of the images are monotone black and white, which shows the same tendency to prioritize nostalgia and historical association. However, the color palette is noticeably more vibrant, with the presence of bright colors. Dominant color combinations include red and white, as well as brown and green. These color choices are significant as they align with traditional Austrian cultural elements. Red and white are commonly associated with the Austrian flag, symbolizing patriotism and national identity. Brown and green hues may represent nature and the Alpine landscape, which are integral to Austrian cultural heritage.

The predominant age representation in these images aligns with the previous datasets, with a focus on older individuals and a dominant male presence. However, there is a noticeable increase in the images of women compared to the person dataset. The inclusion of objects, specifically pictures of traditional Austrian garments such as dirndls and lederhosen, with varying levels of completeness and aesthetic, show the model's commendable attempt. These garments are emblematic of Austrian cultural identity, reflecting specific Bavarian decoration patterns.

Regarding meaning, the Austrian clothing images portray a distinct cultural perspective. The improved quality in images and the inclusion of significant traditional clothes and colors show the model has a fair understanding of Austrian traditional clothing. Similarly, the representation of older individuals in black and white and aged look, emphasizes traditions in the past. Identification of biases or cultural assumptions: the vibrant colors, and the beautiful clothing details show a sophisticated, culturally rich portrayal of Austrian culture.

Overall, it suggests the model's modest understanding of Austrian culture, even though it still lacks authenticity compared to real-life images. The homogeneous pattern of clothing featured in images may reflect stereotypical representations of Austrian in media and tourism, especially in American media, which idolized European culture as "high culture", representing elitism, beauty, and sophistication (Kronenberg, 2007).

Austrian City

Figure 18 Austrian city - Best images

First, compared to the previous human images, the city images are surprisingly high quality, with good contrast, and proportionally correct composition. They also feature a full range of vibrant, accurate colors with the dominance of blue sky, green trees, yellow and white buildings, and turquoise roofs of traditional churches and domes. Even technically difficult details, like typical Austrian-looking church, city layout, and Alpine mountain elements, are accurately represented. The quality of images is high and consistent across the set. The pattern of images is also consistent: most images are in the aerial view, with the common elements including mountains or greenery in the background, European-looking churches and buildings, and a river in the center. The city looks modern but noticeably European-looking with elements of older architecture from the previous centuries. There is no or very minimal presence of modern-day city objects such as high-rise buildings or busy highway roads.

Regarding meaning, the technical and aesthetic quality is high, showing the Austrian city as beautiful and rich in culture. The compositional elements convey a sense of harmony and balance between modern and human life and nature in the scene, implying the long tradition of human-nature harmony in Austrian culture. The images tend to have a wide range of bright colors that show the cities in a blue clear sky and greenery background, creating a vivid and pleasant impression.

Regarding accuracy and biases, overall, the elements in the images are accurate of Austrian architecture and culture. For example, there are turquoise domes and yellow buildings, which were once the popular paint color for their association with the Austrian empire and remained typical with older buildings. The images do not portray any negative stereotypes or biases about Austria or its people. However, the images may overemphasize certain idolized aspects of Austrian cities (such as historical churches and mountains) while ignoring others that are more modern.

Austrian Workplace

Figure 19 Austrian workplace - Best images

Similar to the city images, the workplace images generated are of high quality and consistent across datasets, with correct composition, correct coloring, good contrast, and details. The images look realistic and coherent, without any noticeable artifacts or distortions. All the images show modern office buildings, both in black and white colors, with no visible signs of contextual symbols, such as cultural, national, or location indicators.

I find the generated images have a high modality, as they look realistic and truthful. They also have high intertextuality, as they conform to the genre conventions of realistic photography and evoke associations with Austrian workplaces. However, we also find that the generated images have low ideological diversity, as they reflect a narrow and homogeneous view of Austrian culture and the work environment. The images show only one type of workplace: a modern office building in a city. The images do not represent the variety of the Austrian workplace culture, such as different types of occupations, and industries.

These findings suggest that Stable Diffusion has some cultural biases and stereotypes of Austrian culture in its image generation process. It seems to favor a Western-centric, urban-centric, and corporate-centric view of Austrian work culture. I argue that this lack of contextualization and differentiation may reflect a dominant cultural perspective that assumes a universal and homogeneous notion of the workplace. This may also suggest a stereotype of Austria as a modern and developed country that is indistinguishable from other Western countries.

Overall, the images generated by Stable Diffusion in the context of Austrian culture present a mixed portrayal. The depiction of Austrian culture exhibits some degree of "cultural accuracy" by incorporating Austrian cultural symbols such as red and white dirndl and echoing traditional stereotypes. The aged representation, vintage aesthetic, and nostalgic qualities evoke a connection to the past and historical traditions. The high level of aesthetics, for example, the vibrant city scene, and detailed clothing style, portrays Austrian culture as rich and sophisticated. However, it falls short of capturing the diversity and contemporary of Austrian culture. The images generated are also quite homogeneous, showing a narrow and idealized view of Austrian people, women, and clothing, even though it does fairly well in terms of setting portrayal of the city and workplace. The images predominantly present an antiquated representation, emphasizing a narrow and idealized view of Austrian culture. In the people's images, the dominant male presence perpetuates gender imbalances.

While some elements exhibit cultural accuracy by echoing cultural symbols and traditional stereotypes, the overall representation lacks depth and fails to capture contemporary Austrian culture. In terms of biases and cultural assumptions, the images generated by the model are mostly positive and flattering, with no obvious negative biases portrayed. However, the model fails to capture the diversity and contemporary aspects of Austrian culture, including its gender diversity and modern looks. The overemphasis on aged and formal settings, along with the idealized presentation, further reinforces cultural assumptions of a homogeneous, conservative, and privileged Austrian culture.

Indonesian Culture

Figure 9 Indonesian people image samples

First, before analysis, it is important to note the diverse characteristics of Indonesian people with more than 300 ethnic groups and 700 languages, many of them have a small population and have limited digital presence due to lack of digital access, lack of media presence and documented cultural artifacts. Thus, capturing the representation of Indonesian culture and people are particularly complicated task, one that is challenging for even cultural experts and global archivists. Therefore, there is no single or definitive way to represent the Indonesian people. However, for approximation purposes, some common features that can be used as indicators of cultural identity are:

- Skin color: Indonesian people have various shades of yellow to brown skin, ranging from light to dark.
- Hair color and texture: Indonesian people typically have black or dark brown hair that is straight or wavy.
- Eye shape and color: Indonesian people usually have almond-shaped eyes that are black or dark brown.
- Clothing: Indonesian people wear different types of clothing depending on their region, religion, and occasion. Some traditional clothing items are batik (a patterned fabric dyed with wax), kebaya (a fitted blouse), sarong (a wraparound skirt), songkok/ peci (a cap), and udeng (a headband). Traditional clothing also contains various accessories such as earrings, necklaces, bracelets, rings, scarves, and masks. Some accessories have symbolic or religious meanings, such as keris (a dagger), selendang (a shawl).

The literature review, together with the assistance of an Indonesian acquaintance, helped guide my analysis and selection of the best images.

Indonesian Person

Figure 20 Indonesian person image samples

The Indonesian people's images have good modality in terms of realism, detail, contrast, and color range. Some have more realistic textures and lighting effects to photo-like quality. Images show certain variations in terms of composition, and framing. The images exhibit a higher level of color vibrancy, in both the people's faces, clothing, and accessories. Most of the images are close-up portraits, while those with visible backgrounds suggest rural settings, with traces of bushes and muddy roads. The style and people’s appearance reflect a more modern appearance rather than resembling previous centuries as in the case of the Austrian people.

Regarding the people depicted, there is also a predominance of male individuals in the generated images, but a greater representation of women is also observed. The majority of depicted individuals appear older, with visible wrinkles on their faces. Brown and Dark brown are the dominant skin color. The facial expressions in the images vary between neutral and smiling, conveying a friendly vibe. Gestures or body language are not specifically mentioned in the observations. Apart from the close-up portrait pictures, about half of the images show people seen from the front high angle.

Roughly half of the images show people in casual daywear that is not culturally specific, while the other half depicts individuals wearing typical Indonesian clothing. For example, some men are shown wearing batik-detailed shirts with vibrant colors. Only a handful of images feature women wearing any type of head cover, which is in reality the majority as Indonesia is the largest Muslim nation in the world. The quality of the hijabs appears subpar, characterized by incomplete or shaggy shapes. Regarding diversity, although Indonesia is a multi-ethnic country, the generated images do not exhibit a significant amount of diversity in terms of clothing, hairstyle, skin tone, and other visual cues. The lack of specific ethnic, religious, or cultural symbols makes it challenging to identify the specific ethnicity, religion, or region represented. The overall appearance of the individuals lacks distinct regional or cultural variations.

The visual elements in the generated images combine to reflect a representation of Indonesian people in their everyday lives in rural settings. The vibrant colors and clothing, and natural backgrounds convey a sense of vibrant cultural heritage, albeit the contrast and vibrancy were over-extended.

The images generated by Stable Diffusion have low cultural accuracies. In terms of skin color, hair color and texture, and eye shape, the overall representation bears somewhat resemblance. However, the skin tone is over-contrasted and saturated to reddish brown, while the majority of Indonesian have yellow to medium brown. Also, while it seems like the model attempts to depict people in casual wear, the use of colors, and ornaments in clothing that are not typical for everyday Indonesian clothing. For example, the image of the man wearing a cap and glasses does not match the common image of an Indonesian man, who usually wears a peci (a traditional hat) or no hat at all. For women, there is an over-emphasis on what appears to be the model's attempt to add batik/ elaborate details to their headpieces/ hijab, but the color and details are too bright and contrasted for what people often wear, which usually has more muted and harmonious colors. The way the hijabs are worn is also odd, for example, tugging the hijab inside the shirts, but this can be attributed to the fact that the model needs more steps to complete generating. Interestingly, one argument in favor of the model is that as Indonesians are so diverse, and the way the people are depicted can bear resemblance to some minority ethnic groups in remote areas, however, even so, there is an over-representation of these minorities and a lack of cultural symbols or elements from the majority of other Indonesian cultures, such as Javanese (the largest ethnic group in Indonesia) with batik, kebaya.

The limited representation of cultural symbols and the homogeneity in appearance across the images suggest a lack of diversity and fail to capture the distinct cultural nuances of Indonesian people. The pattern of people being shown from a high angle also has significance. According to Kress and van Leeuwen's visual grammar theory, this angle contextualizes power and involvement. In this case, the people are shown in more subordinate positions than the viewers. Digging deeper, most people are shown in casual, sometimes shaggy attire, with weathered faces with wrinkles, in visible rural settings with some presence of disheveled bamboo walls or dirt in the background. This may convey that the depicted people are not in a modern environment and are not in a well-off condition. When compared to other images of Indonesian people online, this is far from the reality of Indonesian as a lower-middle income country that is rapidly developing and has a growing urban percentage.

Overall, the Indonesian images show a high level of modality, and technical quality and suggest an attempt to represent Indonesian people in a typical daily life appearance with fair accuracy of physical body appearance. However, the skewed gender representation, limited religious, ethnic, and regional diversity, and inaccurate representation of people's appearance and cultural elements greatly undermine its performance.

Regarding bias and cultural assumption, the image in general show Indonesian person in a somewhat positive light with neutral to friendly expression. However, the recurring high angle, combined with the over-representation of rural settings and people in weathered attires, may convey an image of Indonesian people that is subordinate/ distant from the current modern "Western" society. Additionally, the vibrancy in colors, the way most people depicted dress and appear in rural settings, while true to some minority Indonesian culture, is an over-representation and shows a bias to heighten the "differentness" of Indonesian. It is fair to say this corresponds with Orientalism and perpetuates the stereotypes of Indonesian as either exotic foreign culture or an underdeveloped nation. This perspective reflects a Western-centric bias, though inadvertently as the result of the training dataset scarped from Western-dominant sources like online forums and travel blogs. It is also lacking in diversity in terms of ethnicity, religion, region, or class, stripping Indonesia from its rich, diverse cultures and instead providing a specifically flat, inaccurate, skewed representation of certain minority sub-groups in Indonesia.

Indonesian Woman

Figure 21 Indonesian woman image samples

Similar to what was observed in the Indonesian people set, the images of Indonesian women exhibit a higher level of contrast and vibrancy compared to the Western culture set. The individuals depicted typically have dark brown to black eyes, black hair, and dark brown skin tones. Most of the images consist of close-up portrait shots without significant background details. Some images suggest rural settings, similar to the people dataset. Most of the depicted individuals appear to be older, over 30 years old, as indicated by visible wrinkles on their faces.

The representation of Indonesian women in hijabs is more prominent in the generated images compared to the people dataset. This aligns with the cultural reality of Indonesia, where Islam is the majority religion. However, the hijabs depicted in the images often exhibit vibrant colors or intricate details, which may not accurately reflect real-life practices. Additionally, individuals are often seen wearing shawls wrapped around their bodies. The positioning of the hijabs is not always accurate, likely due to the complexity of rendering and the limitations of the model.

The combination of visual elements, such as vibrant colors, intricate clothing details, and rural backgrounds, contributes to a representation of Indonesian women that emphasizes exoticism and a specific group of women in remote areas of ethnic minorities. These elements, though are part of the Indonesian cultural facets, may not capture the full diversity and complexity of the country's population and fail to capture the generalized representation of the majority sub-cultures.

Regarding cultural accuracy, the over-contrast and over-vibrancy of colors and patterns in the clothing worn by Indonesian women similarly suggest a pattern of exoticism. The representation tends to focus on older women in rural settings, perpetuating stereotypes of Indonesia as a backward and rural nation. This aligns with the exoticism and orientalist perspective.

Indonesian Traditional Clothing

Figure 22 Indonesian traditional clothing image samples

Generated Images	**Real-life photos**
 	 
Image generated by Stable Diffusion v1.5	Photo of Indonesian traditional clothing – Bali (left) and Javanese (right) Source: (Febryandini, 2018)

Table 6 Comparison of Indonesian traditional clothing images

The generated images of Indonesian traditional clothing in Stable Diffusion exhibit a higher level of contrast, detail, and vibrant colors compared to other datasets. The dominant colors observed are red and yellow, which are often associated with gold embroidery commonly found in Bali and Padang traditional clothing. The composition of the images shows improved quality compared to previous datasets, with people correctly positioned in the center and without awkward cropping. Most people depicted in the images are male, but there is a more balanced representation of age groups compared to other datasets, including individuals in their 25-40 years old. Facial expressions tend to portray friendliness or smiling, contributing to a positive representation of individuals.

The clothing accessories displayed include intricate headbands, headpieces, elaborate flowery and gold necklaces, and embroidered garments. Some males are depicted with a cap, which resembles peci, a traditional cap, although inaccuracies in shape and color are present. Intricate and elaborate head accessories, more commonly associated with female Sumatrans during special occasions, appear on males in the generated images, suggesting a model's error. The clothing predominantly features vibrant red and yellow colors, which is a significant color in a few Indonesian cultures, however, this is more prominent in Indigenous groups. Whereas the traditional clothing of the majority of cultures, such as Javanese/ Borneo, in reality often incorporates more neutral colors such as white baju koko and black pants in combination.

The combination of visual elements, such as vibrant colors, elaborate accessories, and depictions of individuals in rural settings, perpetuates exoticism and orientalist views. The overemphasis on elaborate accessories and inaccuracies in their attribution to gender contribute to an exoticized representation of Indonesian traditional clothing. Also, most people are pictures with similar dark-brown skin tones, and clothing styles, which can be argued to be the model's attempt to represent certain Indigenous groups in Indonesians. Nevertheless, it reflects a specific cultural perspective that aligns with stereotypes and does not capture the general representation and authenticity of Indonesian culture.

The analysis indicates both poor cultural accuracy and biases in the model's view of Indonesian culture. While efforts are made to represent specific cultural elements, such as traditional clothing and accessories, inaccuracies and overemphasis on exoticism and vibrancy contribute to biased representations. The inaccurate placement of headgear and other accessories can be attributed to both the model's over-bias towards generating male images, and its poor understanding of Indonesian traditional clothing and culture overall.

Indonesian City

Figure 23 Indonesian city - Best images

Similarly, the generated images of Indonesian cities exhibit a higher level of modality than the people images, in terms of contrast, clearness, and range of color. The dominant color observed is turquoise blue, depicting the presence of rivers and the ocean within the cityscape

Most of the images display correct composition, offering aerial shots of the cities from an elevated perspective. The cityscapes portray a mix of modern elements, such as high-rise buildings and highways, along with characteristics commonly found in developing countries, such as low-rise houses with red panel roofs. This combination accurately represents the reality of urban settings in Indonesia. Some images depict typical Southeast Asian streets with messy roads and low-rise houses, further capturing the diverse urban landscape.

While the majority of images do not exhibit visible cultural signs or objects, this can be attributed to the nature of aerial shots that predominantly capture cityscapes and streets from above. A few images (less than 10) depict buildings that resemble temples, suggesting a recognition of local cultural elements by the model. Although limited in number, these images contribute to the understanding of the local cultural context.

The combination of visual elements, such as the vibrant turquoise ocean, green trees, modern infrastructure, and characteristics of developing countries, signifies the dynamic nature of Indonesian cities, where traditional and modern aspects coexist. The absence of specific cultural symbols may be attributed to the chosen perspective of aerial shots, which tend to focus on the overall urban landscape rather than specific cultural details.

The analysis of the generated images of Indonesian cities in Stable Diffusion indicates a relatively fair representation in terms of cultural accuracy. The presence of modern infrastructure and elements of developing countries reflects the reality of Indonesian urban settings. The presence of vibrant colors, though present in Indonesian city space, can be a bit exaggerated and links to exoticism, albeit to a limited extent. This can be attributed to the training dataset which may contain images of Indonesian streets as a travel destination in Western tourism and media.

Indonesian Workplace

Figure 24 Indonesian workplace - Best images

The Indonesian workplace images exhibit good overall quality in terms of contrast, color, and detail. The composition of all the images accurately portrays a workplace setting from an eye-level view. Most of the pictures depict people and tables within an office setting. However, unlike the Western cultural workplace set, there are occasional traces of non-office settings, such as factory floors and factory ceilings. While people are not frequently or entirely pictured in workplace images, the model attempts to incorporate local human objects. The individuals depicted typically have dark hair and colored skin tones, representing the majority look of Indonesian ethnicities.

The combination of visual elements, such as office furniture and occasional factory elements, contributes to the creation of an Indonesian workplace image. The composition and accurate portrayal of workplace settings enhance the cultural accuracy of the images.

The analysis indicates a relatively fair representation in terms of cultural bias as there is no clear prejudice. Even though some of the offices look more like a factory and less modern, and more chaotic than their Western counterparts, this can be argued since Indonesia is a developing country where the other types of offices are more present.

Similar to another set, the model portrays a fixed understanding and does not incorporate other types of jobs, practices, and cultural symbols other than generic white-collar office settings.

Irish Culture

Irish People

Figure 25 Irish people image sample

The generated images primarily exhibit two distinct visual styles: one-third of the images are in black and white or monotone colors, displaying lower quality, while the remaining images are colored. Overall, the images demonstrate a relatively higher modality in terms of contrast and details, particularly when compared to the images of Austrian individuals. However, the quality of the images varies across different groups and individuals, with some showcasing realistic and intricate features, while others appear blurry and caricature-like. Regarding composition and background, the majority of the images depict portraits or close-up headshots with minimal or plain backgrounds. Although only a few images provide visible background information, they often feature natural or rural settings, such as green grass or mountains. Notably, dominant colors in the images include green, orange, and brown, with recurring elements like green shirts, green hats, ginger/orange hair, and beard, as well as brown shirts and backgrounds. The diversity was limited to different groups and individuals. For example, most of the male images had similar facial features such as a beard, mustache, or stubble.

Most of the female images had similar facial features such as long ginger hair. Most of the red-haired images had similar clothing such as green shirts or hats. Most people have similar, neutral expressions, not smiling or serious.

The combination of these visual elements conveys specific meanings. The utilization of black and white colors may evoke a sense of nostalgia, reminiscent of old photographs. The presence of vibrant colors like green and orange, which are strongly associated with Ireland, can be interpreted as symbolic representations of national identity and pride. Additionally, the inclusion of nature and rural settings may connote the importance of the natural landscape and its connection to Irish culture.

All of the images were white, and male, with less than 5 images of women and no representation of other ethnic groups. The most common hair colors were ginger, brown, and other. The most common eye colors were blue and green. In comparison, 2016 statistical data on Irish demographics shows 50.6% are female and 49.4% are male. The majority of people (82.2%) identify as white Irish, followed by other white (9.5%), Asian (2%), and others. The most common hair colors are brown, ginger, and black. On a generalized scale, the ethnic representation of the people shown through skin tone, eye, and hair colors, is justifiable.

Overall, the images demonstrate some adherence to cultural attributes associated with Irish culture, such as the usage of specific colors and references to natural landscapes. However, it is important to note that the diversity within the images appears limited, both in terms of different groups and individuals. The male presence heavily dominates, and there is a lack of representation of various age groups and other demographic characteristics.

Irish Woman

Figure 26 Irish woman image sample

When examining the dataset of Irish women images, similar characteristics can be observed as with the overall dataset of Irish people. The image quality remains inconsistent, with more than half of the dataset being in black and white. The color images, on the other hand, exhibit good modality with adequate contrast and complete composition. Notably, a prominent presence of green clothing, green backgrounds, and orange/ginger hair is observed in the color images, which aligns with cultural associations with Ireland and its symbolism. Similar to the general dataset, the presence of backgrounds is not prevalent, but among those images that do feature backgrounds, green and brown colors are commonly depicted. This suggests that the women are often portrayed in rural or natural settings, further reinforcing connections to Irish landscapes and surroundings.

Also, most people appear to be white Irish, shown through skin tone, hair, and eye color. Furthermore, a significant portion of the depicted women appears to be older, despite no explicit prompt or description indicating their age. The emphasis on older women may imply a cultural association with tradition, or heritage. In terms of visual composition, the majority of the women are shown in posing positions, as if they were being photographed. Facial expressions tend to be predominantly neutral.

Regarding accuracy, ethnic distribution and the presence of cultural symbols and associations such as green colors and rural settings aligns with certain cultural aspects. However, the limited diversity in terms of age representation and the potential bias towards older women. This suggests a partial and somewhat rigid and outdated representation that does not fully capture the nuanced realities of Irish womanhood.

Irish Traditional Clothing

Figure 27 Irish traditional clothing image samples

Compared to the datasets of individuals, the clothing images exhibit a higher overall quality. The images demonstrate improved contrast, composition, and details, providing a more visually diverse range of colors. As with the previous datasets, green remains the dominant color in traditional clothing images. Various clothing items feature shades of green, while orange and ginger hues are prevalent in hair and beard representations. Similar to the people dataset, the majority of the depicted individuals in the clothing images are white males, although women are more prominently represented. It is worth noting that many of the male individuals are portrayed with beards, potentially signifying a cultural association with masculinity or traditional Irish aesthetics. In images with visible backgrounds, the surroundings typically indicate that the individuals are posing against nature backdrops, often featuring rocks and abundant greenery reminiscent of mountains or grassy landscapes. This contextual element reinforces a connection to the natural environment and rural settings, which are significant aspects of Irish cultural heritage.

Regarding the clothing pieces, men are often depicted wearing green caps or tall hats, while dark green and brown cloaks or vests are also commonly featured. Some individuals are seen wearing what appear to be wool or leather jackets, potentially known as mantles. Women, on the other hand, are typically depicted wearing long dresses with shawls. Interestingly, a recurring item is a pleated kilt worn by male individuals, suggesting a distinctive cultural garment. The clothing items often exhibit green details, with brown elements reminiscent of natural materials such as leather or wool.

In terms of cultural accuracy, Stable Diffusion achieves a reasonable representation of Irish traditional clothing. The presence of green as the dominant color, specific clothing items like Irish mantles and pleated kilts, and the incorporation of natural elements align with recognizable cultural symbols associated with Irish tradition. However, it is important to acknowledge the overrepresentation of men and the recurring presence of beards which may perpetuate biases and a narrow perspective associated with Irish masculinity. Zooming further, some images of male people with green tall hats bear resemblance to the Irish mythical Leprechaun, which is part of the Irish cultural imagery. However, the representation of Leprechaun when prompted about Irish traditional clothing is an over-representation of Leprechaun in media portrayal of Irish which relies on clichés and can perpetuate a rigid stereotypic image of Irish culture.

Regarding the model's prompt alignment, a noteworthy aspect when compared to the Austrian clothing dataset, the other Western culture selected is the modern appearance of the individuals. Despite wearing traditional attire, the individuals in the images exude a contemporary aesthetic. This shows the model has an understanding of the "realistic" instruction, but it fails to do so for the Austrian traditional clothing dataset.

Irish City

Figure 28 Irish City - Best images

The city images within the dataset demonstrate a higher level of technical and semantic quality, exhibiting clear, well-composed pictures with a diverse and accurate color range. These images often adopt an aerial viewpoint, providing a birds-eye perspective of the cityscape. Notably, the city images incorporate a blend of modern and traditional elements. Modern features such as high-rise buildings and buses coexist with older buildings and churches, showcasing a typical British/ Irish architectural style. The presence of abundant green elements suggests the inclusion of parks or natural surroundings, and images of rivers contribute to the depiction of urban waterways. In addition to the general European architectural style, the Irish-specific elements of city architecture are also captured in the images. For instance, a bridge over a river is depicted alongside older English-style buildings, closely resembling the actual scenery of Dublin City and the Canal Dock area. The high level of accuracy in incorporating cultural elements specific to Ireland indicates a commendable degree of cultural accuracy in the generated images.

Analyzing the social semiotics meaning, the combination of modern and traditional architectural elements captures the coexistence of modernity and tradition, emphasizing the rich historical background of Irish culture while acknowledging its integration into contemporary urban settings. The inclusion of green elements and rivers signifies the integration of natural environments, highlighting Ireland's association with lush landscapes and water bodies. The inclusion of specific Irish architectural elements further reinforces the cultural specificity of the images, enhancing the representation of Irish cultural identity.

Evaluating the model's cultural accuracy, it can be concluded that Stable Diffusion demonstrates a commendable level of accuracy in representing Irish cultural elements in the generated images. The incorporation of distinct architectural features and the presence of specific cultural markers contribute to a more authentic representation of Irish traditional clothing and the associated urban environment. Regarding biases and cultural assumptions, the generated images do not contain negative imagery. Overall, it suggests a positive, culturally distinct, and modern image of Irish cities.

Irish Workplace

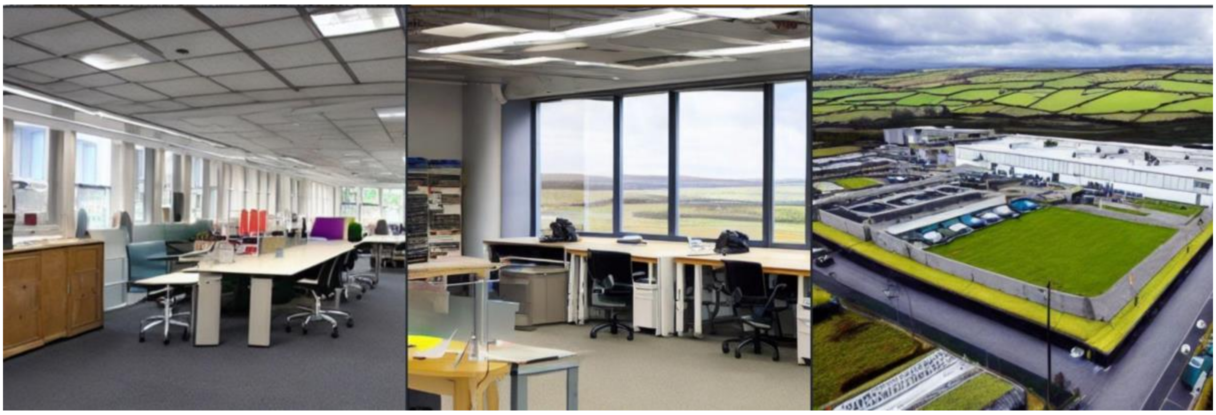


Figure 29 Irish workplace - Best images

The images in this dataset demonstrate good technical quality in terms of color accuracy, contrast, and composition, similar to the observations made for the U.S. and Austrian datasets. Colors are significantly more present and prevalent, with white and green being dominant. Most of the images depict scenes of a typical Western white-collar office environment. Some images present an aerial view of a building complex, characterized by the presence of green spaces that signify a connection with nature. The predominant visual elements include typical office items such as office tables, chairs, and document cabinets, which portray modern white-collar offices, with no discernible signs or symbols specifically associated with Ireland or Irish culture.

Analyzing the social semiotics meaning, the presence of modern white-collar offices reflects the globalized nature of contemporary work environments. The inclusion of aerial views and green spaces may connote notions of spaciousness, tranquility, and harmonious integration of nature within the workplace setting. This may also point to the reputation of Ireland and its green, spectacular nature.

The combination of these visual elements creates an image that reflects a universalized concept of a modern workplace. Though it fails to capture the distinct cultural characteristics of the Irish, it is justifiable since this resonates with the globalized office work cultures dominant in Ireland. No negative stereotype or imagery is detected. Evaluating the model's cultural accuracy, it seems that SD performs fairly well and fairly here. However, similar to another country's dataset, the model seems to have a singular view of what is considered a workplace and does not account for other types of jobs other than office work.

Japanese Culture

Japanese Person

Figure 30 Japanese people image sample

Figure 31 Japanese people - best images

The images vary greatly in their quality and clarity. Some images are sharp and detailed, while others lack contrast and colors. The images were generally fair in terms of modality, but they have less clarity, contrast, and color range compared to the images generated for other countries such as Indonesia. The images are mostly in black and white or monotone, with only one-third in color. The black and white or monotone images had a nostalgic or historical feel, while the color images had a more contemporary or realistic feel. The presence of black and white or monotone images with an old-style appearance may evoke a sense of nostalgia or reference historical Japanese visual styles, such as vintage photographs or traditional Japanese postcards. The background is minimal and mostly plain. The indoor scenes were mostly simple and minimalist, with bamboo materials such as tatami mats and sliding doors.

These are typical elements of traditional Japanese architecture and interior design and reflect the aesthetic principles of simplicity, harmony, and naturalness associated with Japanese culture.

Most images also show people with distinctive facial features that are stereotypically associated with Japanese people, such as almond-shaped eyes, straight black hair, or pale skin. Some images show people with unnatural facial expressions, such as rigid smiles or blank stares. Some images also have unrealistic skin textures, such as smooth and shiny surfaces that resemble plastic dolls. A few images even show cartoon-like or anime-style characters. Only a small subset of images show more realistic-looking people with natural expressions and skin textures. Most images are close-up shots of the person looking directly at the camera. Some images, especially of women, show them looking away or down. The facial expressions and gestures are mostly neutral or calm with rigid posing styles.

Regarding age and gender, the images have more diversity. There is a wider age range feature, with most people being adults from their 20s to 50s or 60s. One notable observation is the higher representation of women in the generated images compared to images depicting other cultures such as Austrian and Indonesian. This gender distribution could be influenced by societal factors and cultural stereotypes that associate women with traditional Japanese aesthetics and clothing.

Regarding clothing, almost all images show people wearing traditional Japanese clothing, such as kimonos and yukatas. This traditional clothing signifies the model’s understanding of Japanese culture and heritage. However, this representation of traditional attire without a specific prompt shows biases in the dataset used to train the AI model or cultural assumptions surrounding Japanese culture and its visual symbols. This may also reflect a cultural bias or stereotype that associates Japan with its traditional aspects more than its contemporary ones.

Overall, the analysis reveals both aspects of cultural accuracy and cultural biases. The generated images provide glimpses of Japanese cultural elements by accurately including elements typical of Japanese people and culture, such as tatami mats, and kimono clothing with cherry blossoms and flower details. The images also show more diverse age and gender representation, especially compared to other cultural images.

However, it is crucial to analyze critically the pattern of certain aesthetics and posing styles. Most people pictured have rigid postures, neutral and serious facial expressions, and formal posing styles in traditional clothing. This may suggest a sense of respect, hierarchy, and collectivism in Japanese culture, as well as a preference for preserving the past and maintaining harmony. This calm expression and rigid posing style characteristic resemble the stereotypical depictions found in traditional Japanese postcards or prints, which often portrayed individuals in a composed and contemplative manner. The combination of these visual elements creates an image of Japanese culture that is traditional, refined, and serene.

Drawing from the theoretical frameworks, these images reflect a cultural bias and stereotype that exoticizes or essentializes Japanese culture as homogeneous, static, and isolated from the modern world. We can see some traces of Orientalism in the representation of Japanese people as traditional, conservative, and formal. Also, looking at intersectionality, the tendency for individuals, particularly women, to look away or down in the images can be interpreted in different ways. While it may reflect a cultural norm of modesty or humility, it also confirms the dominant ideas of East Asian women in Orientalism as obedience or submission. This is a common discourse in Orientalism which tends to fetishize or romanticize the East as a source of mystery, beauty, and spirituality (Hall, 1997). This is likely due to the Western bias in training data, which the AI model may draw on and result in the exoticized representations of Japanese people, especially women, that emphasize their difference and otherness from Western norms.

Japanese Woman

Figure 32 Japanese women image sample

Figure 33 Japanese women - Best images

The images generated for Japanese women exhibit a similar level of detail and clarity compared to the images of Japanese people. Slight improvement in contrast and color allows for a more accurate portrayal. First, similar to the images of Japanese people, the majority of images depicting Japanese women are in black and white or monotone. Similar to the people's images. the texture, posing styles, and color grading in some images resemble postcards or old photographs from the 18th century. However, some of them also have a stylized or artistic quality, such as the ones that look like traditional Japanese drawings. This may suggest that the model is influenced by popular historical paintings of Japanese women, rather than by real people and modern photographs.

Regarding composition and Poses, the images predominantly feature portrait or half-body shots of Japanese women. While they maintain direct eye contact with the viewer, there is a noticeable tendency for some women to look away, reflecting a pattern observed in the images of Japanese people. This pose composition can convey a sense of introspection or modesty. The background in the generated images is minimal, often depicted with plain colors. However, the presence of brownish tones and textured surfaces suggests wooden or bamboo walls. Some images provide additional details, such as the depiction of tatami mats or wooden floors, suggesting traditional Japanese settings.

In the few images with color, red, pink, and yellow are the dominant colors in the clothing worn by the Japanese women. These colors hold cultural significance in Japanese tradition and are popular in traditional clothes, with red symbolizing good luck, pink representing cherry blossoms, and yellow symbolizing prosperity.

Regarding people features, the representation of Japanese women in the generated images displays a relatively homogeneous set of features, including black hair, dark brown or black eyes, fair yellow skin, small almond-shaped eyes, and thin eyebrows. Most women are adults in their 20s, 30s to 50s. The age range is more diverse, and younger, compared to another woman dataset.

Similar to the people images, even without being told to, all the generated images depict Japanese women wearing traditional clothing, such as kimonos and yukatas. These garments reflect cultural authenticity, highlighting the richness of Japanese fashion traditions. Additionally, the presence of hair accessories, such as hairpins and well-styled buns adorned with floral ornaments, aligns with traditional Japanese hairstyling practices in the Edo era.

In terms of meaning, the use of black and white or monotone color schemes, textured backgrounds, and stylized poses evokes a sense of historical significance and nostalgia. These visual elements may contribute to the representation of Japanese culture as timeless and deeply rooted in tradition.

Regarding cultural accuracy, the visual elements and people features observed, such as black hair, bamboo materials, and traditional Japanese clothing, align with general attributes and popular motifs associated with Japanese women. This shows the model has an understanding of Japanese culture to some extent.

However, it is crucial to examine the overplay of some elements that show cultural biases and stereotypes about Japanese women. The overrepresentation of traditional clothing, such as kimonos and yukatas, without any inclusion of modern attire, perpetuates the stereotype of Japan as a static, conservative culture and shows the model heavily relies on historical Japanese images. Furthermore, the stylized poses, doll-like facial features, and nostalgic aesthetics resembling postcards or old photographs reflect Orientalist tendencies.

This is even more noticeable when compared with how the model depicts women in other cultures, for example, Indonesian and Irish, where it mainly portrays older women with wrinkled skins and rural settings. In contrast, there is a strong pattern to include the aesthetic portrayal of Japanese women. These visual elements confirm the exoticization of Japanese women and reinforce the perception of Japan as otherized, idealized, and exoticized. Furthermore, a pattern that persists in the Japanese people's images is a woman looking away or looking down from the viewer. This angle, together with the stiff and reserved facial expressions reinforces the Orientalist stereotypes portraying Asian women as submissive or obedient.

Japanese Traditional Clothing

Figure 34 Japanese traditional clothing image sample

The images generated for people in traditional Clothes exhibit a significant improvement in modality compared to the images of Japanese people. They display enhanced color representation, greater contrast, and higher levels of detail. The dominant color in these images is red, appearing in almost all pictures. It is often accompanied by white and some blue. The use of vibrant colors reflects the traditional Japanese aesthetic, with red symbolizing good fortune, joy, and celebration in Japanese culture.

The background in these images is more visible and detailed compared to the images of Japanese people. Some depict plain-colored or textured walls, resembling bamboo doors, while others show individuals photographed outside with lush green bonsai trees or temple grounds. These backgrounds evoke a sense of cultural context and reinforce the traditional Japanese setting.

Most of the images capture individuals from the waist up or from the top to the ankles. However, it is worth noting that some pictures have individuals with their heads cut off within the frame, which may be considered a composition flaw. The majority of people portrayed are wearing various types of kimonos and yukata, with intricate details of Japanese traditional patterns such as cherry blossoms and bonsai trees. Additionally, women are depicted wearing hairpins adorned with flowers, earrings, and holding fans with intricate designs. These objects and clothing choices serve as cultural symbols that reflect traditional Japanese aesthetics and customs.

The depicted individuals in the generated images range from young adults to middle-aged individuals, typically between 20 and 50 years old. Although there is gender diversity, women are more prominently featured. Interestingly, some images showcase men wearing kimono with female-style kimono and accessories, which is likely due to the model's mistake and over-representation of Japanese women in traditional clothing in the training dataset.

Similar to the previous set, all people have black hair, small almond-shaped eyes, with fair yellow skin tone. Most individuals in these images adopt a rigid and formal posture, often crossing their hands together or positioning them in the middle of their bodies, as commonly seen in traditional kimono posing. Facial expressions are predominantly neutral or reserved, similar to people's images. Interestingly, compared to the previous set, the clothing images have people appearing more relaxed, making it more human-like. However, as a whole, the generated images exhibit reserved, restricted expressions and rigid posing. In the context of wearing a kimono, individuals may adopt a more formal and reserved demeanor.

Overall, the images generated for traditional clothes exhibit better modality than the previous set, demonstrating improvements in color representation, contrast, and detail. However, some images suffer from technical flaws, such as heads being cut off within the frame. In terms of cultural accuracy, the images successfully capture certain elements of traditional Japanese attire and aesthetics. The presence of various types of kimonos and yukata, adorned with intricate patterns and designs, reflects the rich textile heritage of Japan. The incorporation of cultural symbols such as cherry blossoms, bonsai trees, and hair accessories like hairpins further reinforces the cultural context. While this is not the real representation of the diverse Japanese clothes style and culture, it is a commendable attempt at a generalized perspective.

On the other hand, similar to the people images, the analysis of these generated images raises questions regarding potential cultural biases and stereotypes.

The nature of the kimono, with its defined silhouette and restrictive nature, may naturally influence the way people stand and move and make people look more rigid and doll-like.

However, the depiction of individuals with reserved posing can play into the stereotypical Japanese portrayal in Western media and tourism brochures that inadvertently reinforce stereotypical representations of Japanese people as stoic, unemotional, or submissive. It also contributes to the exoticization of Japanese culture and perpetuates narrow and one-dimensional views of Japanese culture. The overrepresentation of women in kimono, with homogeneous aesthetic appearances with a very particular hairstyle, makeup style, and skin tone, can be seen as conforming to Western beauty standards and the objectification of women. This pattern can be attributed to the Western bias in training data and Western fascination with Japanese aesthetics and reinforcing preconceived notions of Japanese culture.

Japanese City

Figure 35 Japanese city - Best images

The modality of the images generated for Japanese cities is generally good to high quality, with clear details and varying levels of visual quality. The color palette in these images is diverse, ranging from vibrant and colorful to more muted and neutral tones. Dominant colors include shades of red, brown, green, and gray. Some images also feature traditional Japanese colors such as vermilion and indigo.

The composition of the images varies, with some focusing on city landscapes, while others highlight specific architectural elements or cultural landmarks. There is a mix of close-up shots and wider angles capturing the cityscape.

Many images depict traditional Japanese architectural elements, such as pagodas, torii gates, shrines, and temples. There are also modern buildings and skyscrapers, showcasing the blend of traditional and contemporary architecture in Japanese cities. The presence of well-known landmarks like Tokyo Tower and Mount Fuji is also noticeable. Several images show bustling streets, pedestrian crossings, and urban infrastructure like train stations and roads. These images convey a sense of urban liveliness and the efficient transportation systems present in Japanese cities Some images include natural elements such as cherry blossoms, parks, gardens, and rivers. These elements highlight the harmonious integration of nature within Japanese cities. While there are some images featuring people, they are often small in scale and not the focus. Individuals depicted are typically dressed in modern attire, such as business suits or casual clothing, reflecting the contemporary urban lifestyle in Japan.

Traditional cultural symbols like lanterns, Japanese characters (kanji), and traditional signage can be seen in several images. These symbols add cultural depth and context to the representation of Japanese cities. There is a noticeable blend of both traditional and modern in the city space, for example, Tokyo-style streets laid with neon signs together with red lanterns and traditional temples.

In conclusion, the images generated for Japanese cities generally exhibit a high level of visual quality and capture various aspects of Japanese urban life and architecture. They showcase the blend of traditional and modern elements, reflecting the unique character of Japanese cities. The inclusion of cultural landmarks, architectural features, and natural elements contributes to the representation of Japanese culture and aesthetics.

In terms of cultural accuracy, the images generally align with common visual representations of Japanese cities, featuring iconic elements that are often associated with Japan. There is no negative bias detectable. Overall, these images portray Japanese cities as beautiful and culturally rich, having a mixture of both traditional and modern life and distinct Japanese cultural symbols. Zooming out, this can be attributed to the growing popularity of Japanese culture among the general global public as both an advanced country and a popular travel destination. While this can play into the idealized notions of Japan as a “utopia” country, on a general scale, it is justifiable to say that the model has an accurate and positive representation of Japanese cities.

Japanese Workplace

Figure 36 Japanese workplace - Best images

The modality of the images generated for Japanese workplaces is generally good, with clear details and varying levels of visual quality. The color palette in these images is diverse, ranging from vibrant and saturated colors to more muted and neutral tones. Common colors include shades of white, black, gray, and various hues of brown. The composition of the images varies, with some focusing on office interiors, while others showcase individuals or groups of people engaged in general office activities. There is a mix of close-up shots highlighting office details and wider angles capturing the overall workplace environment.

Regarding workplace objects, all images depict modern office settings with desks, chairs, computers, and other typical office equipment. The office spaces vary in size and style, including open-plan layouts, cubicles, and conference rooms. Some images also showcase traditional Japanese elements such as sliding doors (fusuma) or tatami mats in designated areas.

The images feature individuals and groups of people engaged in various general office activities, including individuals working on computers, writing, or participating in meetings. Most individuals are depicted in professional attire, such as white shirts and business suits, suggesting a formal work environment. Both men and women are represented, though there may be variations in gender distribution across different images. Gestures and facial expressions are not clearly visible but reflect a uniform, professional demeanor. Most people are shown working independently or looking at computers. The depiction aligns with the prevalent work culture in Japan, which values diligence and individual responsibility.

While the primary focus is on the workplace environment, some images include subtle cultural elements. These may include bamboo doors, wooden floors, traditional artwork, or decorative items reflecting Japanese aesthetics. However, the presence of such elements may vary, and they are not as prominent as in images specifically focused on cultural representations.

The images generated for Japanese workplaces generally demonstrate a good level of visual quality and depict various aspects of modern office environments. They showcase the presence of technology, formal attire, and typical work activities, aligning with common representations of work settings in Japan. The inclusion of traditional elements, although less prominent, adds cultural depth to the portrayal of Japanese workplaces.

The images generally present a culturally accurate representation of Japanese workplaces, featuring recognizable elements such as office settings, attire, and work-related activities. However, there is a noticeable similarity in appearance among the individuals depicted in the images. While variations in clothing style or position may exist, there is a tendency towards a uniform look, which can be interpreted as individuals wearing similar attire or adhering to a professional dress code. This uniformity may reflect the reality of formal and neutral dress expectations in Japanese workplaces. However, it may overplay the traditional "worker-bees" stereotype that Japanese are industrious and conform to uniformity and does not reflect the modern, diverse work environment in modern Japan.

Pakistani Culture

Pakistan Person

Figure 37 Pakistani person image sample

Figure 38 Pakistan person - Best images

Overall, the images generated have a fair modality, meaning that they are relatively realistic and naturalistic, but not as sharp or detailed as some other images, e.g. Indonesian and Vietnamese people. The images overall have clear central subjects and realistic colors. Contrast and lighting are balanced, although not as sharp as Southeast Asian culture images. The images have a medium to low contrast and a muted color palette, with dominant hues of green and brown, albeit not in bright shades. Many individuals are depicted wearing greyish-green shirts, and there is a notable presence of clothing resembling the Pakistani green flag.

Brown is prevalent in the background, often portraying muddy roads, wooden debris, and sandy environments. Skin tones range from brown to dark brown.

Most individuals are captured in a half-body crop, while others are shown in full-body or close-to-full-body shots. Unlike images from other cultures, Pakistani individuals are often depicted from a farther distance, showing a full body or half body. The background is notably more visible and has more context compared to images from other cultures. It often includes elements that evoke a rural setting, capturing individuals in candid moments walking on a muddy road or swatting down on a soil ground.

The ethnic features are most consistent with the South Asian region, such as dark hair, dark eyes, and brown skin tone. The depicted age range varies, with the majority falling within the adult category (30 to 60 years old). The gender representation is overwhelmingly male, with women comprising a small fraction of the images. Most men are seen with similar facial hair styles, and wearing headscarves or turbans, which are often used by Muslims. Meanwhile, the clothing styles vary between casual plain shirts in grey or greyish green and more traditional Pakistani clothing such as shalwar kameez, and kurt. While women are virtually invisible, of the few that portray women, the women are fully covered with hijabs/ scarves and dresses. This style resembles rural Pakistani attire, which together with the background, shows a heavy bias to showcase only Pakistani people in very traditional and rural settings. This observation aligns with my consultations with an actual Pakistani acquaintance.

Approximately two-thirds of the individuals are looking away or sideways, with many exhibiting a contemplative or distant gaze. The expressions tend to be less friendly or joyful compared to Oriental cultures, and their gestures are mostly static or passive. They are either cropped at the waist or shown in full body, but rarely in close-up shots, while this was the norm in other cultural images. Unlike other cultures (Austrian, Japanese, etc.) the people are not in formal portrait photo poses. They look more like caught off guard or candidly photographed on the street.

Using the social semiotics framework, we can infer some meanings of the elements and their combination in the images of Pakistani people. The frequent appearance of the Pakistani flag in the images suggests that the model associates Pakistani people with their national identity and patriotism. Green is often seen as a color of life, growth, harmony, and Islam, which is the main religion in Pakistan. This, together with the prevalence of Islam turban/ hijabs, show a strong emphasis on the Muslim identity in Pakistani people. Meanwhile, the muddy brown color is often seen as a color of earth, dirt, poverty, and backwardness. The prevalence of brown shown in the muddy road, soil, or sandy background, portrays Pakistani people as either living in rural areas or struggling with poverty and underdevelopment.

Regarding composition, compared to other cultural images, the blurry background in the images creates a sense of distance and detachment between the viewer and the scene. The gaze of the people is an important element that indicates their attitude or relationship to the viewer. Most of the people have an averted gaze, meaning that they are not engaging with the viewer directly. This may suggest distance or detachment from the viewer's culture or perspective. Some of them have a direct gaze, which may suggest a questioning look to the viewers as outsiders. The higher angle may also suggest the dominance or superiority of the viewer over the people.

The facial expressions and gestures of the people in the images are also important elements that convey their emotions and attitudes towards the viewer and themselves. Many people have neutral or serious facial expressions which suggests they have less engagement and rapport with the viewer, especially compared to other Oriental cultures who appear more friendly. The people who have static or passive gestures have less agency and power than those who have dynamic or active gestures. Here, the AI model seems to depict Pakistani people as either indifferent or unhappy with their situation.

The analysis above shows that there are strong cultural biases and stereotypes of Pakistani people. According to Orientalism, colonialism has created a binary opposition between the West and the East, or the Occident and the Orient, where the East is seen as inferior, irrational, exotic, and traditional. Compared to other Oriental cultures, the AI model seems to reproduce a stronger, more negative binary opposition in its images of Pakistani people, as opposed to the West. Also, the generated images portray people with similar hairstyle, clothing style, and overwhelming male in rural settings, which further suggest there is a strong association between the model's understanding of Pakistani people and certain group and characteristics.

There are different layers to this association. One, the depiction of rural settings, rumpled clothing, and muddy roads perpetuates stereotypes of underdevelopment, poverty, and lack of education. These representations align with orientalist tropes that reinforce a hierarchical view of cultures, positioning the Western gaze as superior. Further, the portrayal of Pakistani people in predominantly rural settings and their uniform appearance can be seen as reflective of the post-colonial power dynamics. These images may perpetuate the legacy of colonial discourses that marginalized colonized populations, reinforcing notions of cultural uniformity and subjugation. This homogenized representation overlooks the diversity and individuality within the Pakistani population, reinforcing a monolithic view of the culture. The prevalence of Muslim cultural symbols, such as the Pakistani flag, hijabs, and turbans, can be seen as an effort to emphasize the religious identity of Pakistani individuals. However, the limited visibility of women, their confinement to fully covered clothing, and their marginalization to the background reinforce prejudiced notions of Muslim women's oppressed roles and restrict their agency. Also, how the images depicting Pakistani with somber or unfriendly facial expressions can fuel a negative perception of Pakistani society as inhospitable or despondent, struggling with poverty and underdevelopment. Such portrayal, when juxtaposed with prevalent global anti-Muslim prejudices, reinforces negative stereotypes of Pakistani people.

Pakistan Woman

Figure 39 Pakistani women image sample

Figure 40 Pakistani woman - Best images

The images generated for the prompt "Pakistani woman" exhibit surprisingly better quality than the Pakistani people's images. Regarding modality, the images display a realistic quality, with clear details and colors. The lighting and contrast are well-balanced, contributing to a visually pleasing aesthetic. The predominant colors in the images are warm tones, including shades of brown, beige, and red hues. The skin tone ranges from light to dark brown. There is a noticeably wide range of vibrant colors in their clothing, from bright red, and pink to green and blue, reflecting the diversity and vibrancy of Pakistani fashion.

Most women are depicted with light to dark brown skin tones, which correlates with the dominant skin tone in Pakistan. The age range of the women in the images is varied, representing different generations. However, the focus is primarily on young adult women in their 20s to 30s, with limited representation of younger or older individuals. This representation may reflect the emphasis on showcasing the vibrancy and cultural contributions and the underrepresentation of other age groups in the training data.

The composition of the images typically centers on a single female figure, capturing close-up shots that emphasize facial features and expressions. This composition choice deviates from the previous observation of full or half-body shots in the people's images, potentially highlighting the importance of portraying individuality and emphasizing the subjects' identities. The backgrounds in the images are often plain and lack discernible contextual details, suggesting indoor settings. This minimalistic approach directs the viewer's attention to the central subject, i.e. the women and their cultural attributes.

The women's facial expressions vary, ranging from neutral to serene or confident. Some exhibit a slight smile, while others maintain a composed and serious demeanor. Some women look away or gaze directly at the viewers. The gazes in these images are not as intense as those in the people's images, creating a sense that the viewer is an outsider looking into the subjects' world. However, the gaze here overall is less intense and more curious than the gaze of Pakistani people.

Regarding clothing and accessories, the women are depicted wearing traditional Pakistani attire, such as shalwar kameez or sarees, adorned with intricate patterns, embroidery, and vibrant colors. These elements accurately represent the cultural aesthetics and traditions of Pakistani fashion. Many of the women in the images wear head coverings, such as hijabs or dupattas, which are common in Pakistani culture. The variations in the style and manner of wearing head coverings showcase the diversity and personal choices within the cultural practices.

Together, the visual elements in these images combine to reflect the cultural perspectives of Pakistani women, showcasing the aesthetics, traditions, and religious practices within the Muslim community. The vibrant colors, intricate designs, and cultural clothing serve as symbols of rich cultural pride and identity.

In conclusion, the images of Pakistani women in Stable Diffusion demonstrate a notable improvement in terms of image quality, cultural accuracy, and portrayal compared to the images of Pakistani people. The depiction of women in a better modality and more positive light may be attributed to a bias in the model’s training data. Stable Diffusion might have been trained on a dataset that contains images of women with a more updated and accurate representation of Pakistani women. Judging from the emphasis on aesthetically pleasing portrayals of women in traditional clothing, one can guess that the model has been trained using pictures of women for beauty or travel context, e.g. female clothing stores, and travel posters.

However, it is important to analyze the potential underlying biases. The stark difference in portrayal between Pakistani women and men within the generated images indicates the presence of gender biases in the model's training data. Looking at the images generated, while most women are depicted in either submissive or aesthetically pleasing light, men are often associated with either unfriendliness or aggressiveness. These gender biases may reflect societal expectations and stereotypes regarding women's roles as objects of beauty, as well as negative prejudices about Pakistani men.

Furthermore, the emphasis on traditional attire and head coverings and the absence of women in modern clothes showcase a homogenous, traditional portrayal of Pakistani women. This, juxtaposed with the dominant view of Muslims from the outside world, may perpetuate essentialized notions of Muslim identity and reinforce stereotypes about Pakistani women as solely conforming to traditional Islamic cultural attires and norms. The emphasis on traditional attire, intricate patterns, and vibrant colors also aligns with the Western imagination of the "exotic" East, reinforcing Orientalist tropes.

Pakistani Traditional Clothing

Figure 41 Pakistani traditional clothing sample

The images generated in Stable Diffusion depicting a Pakistani person in traditional clothing exhibit an improved level of cultural accuracy compared to Pakistani person images. The modality is realistic, capturing details of color, shape, composition, facial expression, and cultural symbols. The predominant colors within the images are warm tones, including shades of brown, green, and red hues. These colors are often associated with traditional Pakistani clothing.

Regarding gender distribution, the dominant people pictured are women, with only less than 10 percent of men. The women are mostly wearing shalwar kameez, loose-fitting tunics, and trousers, with a dupatta, and a long scarf, draped over their heads and shoulders. The men are mostly wearing a kurta, a long shirt, and shalwar, loose pants, with some also wearing waistcoats or headgear such as turbans. The clothing is colorful and embroidered with various patterns and motifs. The traditional attire worn by the individual represents cultural aesthetics and traditions. The clothing features intricate patterns, embroidery, and vibrant colors, showcasing the rich textile heritage and artistic craftsmanship of Pakistan.

The individuals are typically captured in full body shots, allowing for a broader view that includes details of their attire, physical gestures, and the surrounding environment. The backgrounds are somewhat visible, showcasing elements that suggest the individuals are posing in front of intricate embroidery rugs, decorated mosque walls, or local traditional markets. These details contribute to the overall aesthetic and reinforce the cultural context.

The facial expression of the depicted individual varies, ranging from neutral to serene or slightly smiling. Most people are pictured with a composed and relaxed demeanor. Most people either look away or look directly at the viewers, creating a sense of engagement or introspection. These facial expressions and gestures portray a diverse range of emotions and engagement in a neutral or positive light.

Analyzing the combination of visual elements, these images reflect a specific cultural perspective that somewhat aligns with the majority of Pakistani traditions and values. The accurate portrayal of traditional clothing, intricate patterns, and vibrant colors consolidate a distinct, vibrant depiction of Pakistani culture and conveys a sense of pride in Pakistani culture and traditions. The full-body coverage of both genders also reflects some aspects of Pakistani culture, such as the importance of modesty, especially for women, and the diversity of ethnic and regional influences on the embroidery styles.

However, the images also exhibit some cultural biases and stereotypes, such as the overemphasis on a specific style of traditional Pakistani Muslim clothing and the lack of the presence of other popular clothing items in the Northern regions, such as saris for women or Jinnah Cap for men. The images may imply associating Pakistani traditional clothing with only the Muslim majority, which does not fully capture the diversity of its culture. Furthermore, it is important to examine the overrepresentation of women in traditional clothing critically. The emphasis on women in vibrant, intricately decorated clothing also conforms to the Oriental view of objectifying Asian women and reinforces stereotypes about the otherness of the "exotic East", often in a romanticized or fetishized way.

Pakistan City

Figure 42 Pakistan City - Best images

The modality of Pakistani city image is overall good, with a fair level of contrast, color range, and composition. All images focus on capturing the cityscape in its entirety from an aerial view, depicting a similar composition and perspective. There is a lack of street-level or human-view images, which could suggest a limitation in the AI model's training data.

The color palette consists of earthy tones, including shades of brown, beige, and gray, reflecting the natural and architectural elements of the city. The buildings are typically portrayed in a mix of vibrant and muted colors, representing the diverse architectural styles found in Pakistani cities. The presence of in-city water bodies, such as lakes, adds a touch of freshness. Regarding context, most pictures are overlaid with some blurry or smoky beige-brown background, reminiscent of a sandy or dusty environment typically associated with the Middle Eastern desert. It is worth noting that while some regions in Pakistan, such as Balochistan, feature arid landscapes, the sandy desert aesthetic might not be representative of the entire country.

The buildings are often tall and densely packed, creating a sense of urban density. The images are characterized by dynamic colors, particularly on the roofs of the buildings. However, it is essential to note that in reality, the predominant color palette of buildings in many Pakistani cities is often paler, with yellowish tones, rather than the colorful roofs depicted in the generated images.

The visual elements in these images, such as high-rise buildings and highway roads, reflect features commonly associated with urban landscapes. Together, they combine to represent a modern and evolving Pakistan, showcasing the urbanization, infrastructure, and progress of the country. The vibrant colored elements collectively convey a sense of modernity and dynamism associated with urban Pakistani cities.

However, the absence of clear cultural or religious architectural elements may contribute to a generic representation that lacks specific cultural markers. This absence might result from the AI model's training data, which might not have included sufficient samples or variations of Pakistani architecture.

Upon further analysis, the images depicting a Pakistani city in Stable Diffusion exhibit a fair level of cultural accuracy and a positive representation of the Pakistani urban landscape. However, this is in contrast to the images generated with the prompt focusing on Pakistani people, which predominantly showcased rural settings. The emphasis on high-rise buildings, highway roads, and an aerial view perspective conforms to a Western-centric understanding of cities, which might stem from the biases embedded within the training data. If the training data predominantly includes images of developed urban areas or follows Western aesthetic preferences, the model may learn to prioritize and reproduce these representations in its outputs.

Pakistan Workplace

Figure 43 Pakistan workplace - Best images

The images depicting a Pakistani workplace exhibit a good level of modality and overall cultural accuracy. The scenes predominantly depict a typical office setting, showcasing elements commonly associated with modern office environments such as desks, chairs, and computers. Although the number of people in the images is limited, those depicted have brown skin tones and are dressed in either plain collared shirts or long attire, which is a typical work look in Pakistan.

In terms of visual grammar, the images adhere to principles of composition and representation commonly found in office photography. The arrangement of objects, such as desks, chairs, and office equipment, follows established conventions of an office workspace. The lighting and color balance in the images contributes to a realistic and visually appealing aesthetic.

From a social semiotics perspective, the composition and arrangement of objects within the office environment communicate a sense of order, structure, and routine. The combination of brown skin tones and individuals dressed in long attire suggests a connection to Pakistani cultural norms and practices. The lack of clear portrayal of Muslim clothing or Pakistani-specific elements within the workplace images stands in contrast to the images generated for the prompt related to Pakistani people, where the Pakistani flags and Muslim clothing dominated. This absence may indicate an attempt to portray a secular or neutral work environment, focusing primarily on the professional aspects rather than religious or cultural affiliations. This could be viewed positively as it avoids essentializing or reducing Pakistanis solely to their religious identities.

One notable observation is the absence of women in these images, with only men being depicted. This gender disparity raises questions about gender representation and potential biases within the AI model's training data. The underrepresentation of women in workplace images may perpetuate stereotypes or reinforce societal norms that restrict Pakistani women’s presence in professional settings.

In conclusion, the images generated by Stable Diffusion depicting a Pakistani workplace exhibit good modality and cultural accuracy. The visual elements reflect a typical office setting and adhere to established visual grammar principles. However, the absence of women in workplace images may reflect prevalent gender biases or unequal representation in professional domains. This could potentially perpetuate stereotypes that reinforce existing gender disparities in Pakistan.

Vietnamese Culture

Vietnamese Person

Figure 44 Vietnamese people image sample


Figure 45 Vietnamese person - Best images

The generated images depicting Vietnamese people exhibit good modality, with good contrast, color range, composition, and level of detail. Most images show individuals in a headshot or half-body positions. While the backgrounds lack clarity, they often display elements of green (trees, bushes) and brown (dirt, muddy roads, trees), indicating a rural setting. These visual elements align with the Vietnamese landscape, particularly in rural areas. The color palette in the generated images varies widely, featuring shades of yellow, brown, and red as dominant colors. Although the colors are generally less vibrant compared to Indonesian cultural image sets. The backgrounds of the images, while only partially visible, often include rural elements such as trees, bushes, dirt, muddy roads, and scattered buildings. Some images feature individuals engaged in specific activities, such as riding motorbikes, using wagons, or squatting on the ground. These activities are common, however, together with the background elements, further reinforce the rural context.

Regarding the human subjects, the facial features of the depicted individuals generally include dark brown or black eyes and black hair, which are not entirely inaccurate but may overly emphasize smaller eyes. The individuals typically exhibit friendly or neutral facial expressions, fostering a sense of approachability. The skin tones portrayed in the images encompass various shades of dark yellow and brown, which, while not entirely inaccurate, maybe slightly exaggerated towards minority groups in remote areas, compared to the average Vietnamese. Similar to other cultures, most people depicted are older, above 40 years old with skin with visible wrinkles and weathered folds, further reinforcing their age and rural look.

Regarding clothing, most people are pictured with some type of hat, which can be pointed to the models' attempt to depict the Vietnamese traditional conical hat (non la), however, there is an overrepresentation. The clothing worn by the individuals tends to be casual to very casual, with a range of colors from muted to vibrant. Notably, there is a lack of depictions of modern shirts or traditional attire, and the emphasis is biased toward showcasing rustic clothing styles.

The combination of visual elements, facial features, attire, and expressions shows an attempt to incorporate Vietnamese individuals within their cultural context. However, the focus on rural settings, exaggerated hats, and limited variation in facial features may inadvertently perpetuate certain cultural stereotypes or reinforce Orientalist notions associated with Vietnamese culture.

The analysis of the generated images in Stable Diffusion indicates a low to moderate level of cultural accuracy. While some visual elements align with Vietnamese cultures, such as composition, color, and attire, there are notable inaccuracies and biases present. The emphasis on rural settings, exaggerated conical hats, and limited representation of the younger population, modern clothing, or traditional attire may perpetuate the perception of Vietnam as a predominantly rural, agrarian society. This portrayal overlooks the urban and modern aspects of Vietnamese culture and reinforces Orientalist stereotypes of exoticism and rural primitivism.

Moreover, the overemphasis on specific facial features, such as smaller eyes, and the color exaggeration of skin tones may inadvertently reinforce certain stereotypes associated with South/ East Asian cultures as exotic. These representations largely ignore contemporary culture and show limited diversity within Vietnamese society.

Vietnamese woman

Figure 46 Vietnamese woman image samples

Figure 47 Vietnamese women - Best images

Regarding visual Elements and Composition, The woman images exhibit good modality, with contrast and composition, with close-up headshots or half-body shots of Vietnamese women. The backgrounds, though not always clear, often display traces of green (trees, bushes) and brown (dirt, muddy roads, trees), which convey a rural setting. The images feature a diverse range of colors, with yellow, pink, and blue being dominant. Compared to the Indonesian woman images, the colors in the Vietnamese woman images are less vibrant and more realistic. The clothing colors are varied and realistic, avoiding over-bias towards any specific hues.

For the human subjects, the facial features of the depicted Vietnamese women typically include dark brown or black eyes and black hair. While these features are not entirely inaccurate, there is a focus on smaller eyes, which may reinforce a specific stereotypical East Asian look. The facial expressions of the women are predominantly smiling or neutral, conveying a friendly demeanor. There are varying shades of skin tone, from light to dark yellow skin, which is generally realistic. Notably, the generated woman images tend to have lighter skin tones compared to the generated people images, which may reflect Orientalist notions of lighter skin being preferred or associated with Vietnamese feminism and beauty. Additionally, the representation includes a broader age range, with younger women in their 20s, 30s, and older women in their 40s and 50s.

The clothing worn by the Vietnamese women in the images ranges from casual to very casual attire. The colors of the clothing are muted yet realistic, with a lesser emphasis on vibrancy compared to the Indonesian woman images. However, there is a notable absence of both modern shirts and traditional clothing styles in the generated images. Some images depict women wearing shirts with a "cross collar" style, which is more commonly associated with Chinese clothing (such as the Hanfu jiāo lǐng fúzhuāng). This inaccurate portrayal reflects a misrepresentation of Vietnamese fashion and mistakenly associates it with Chinese culture. Similar to the “people” images, the presence of the conical hat, a traditional Vietnamese, is common in the generated images. While the inclusion of this cultural symbol is commendable, the shape of the hats often appears incomplete and resembles Chinese hat styles rather than authentic Vietnamese conical hats.

The analysis of the generated images of Vietnamese women in Stable Diffusion reveals certain cultural biases and stereotypes. The focus on a rural setting, the overemphasis on smaller eyes, and the misrepresentation of clothing styles and the conical hat shape again perpetuate stereotypes associated with Vietnamese as dominantly rural, agrarian, and not modern. This inaccuracy in depicting Vietnamese clothing reflects a misrepresentation and a conflation of Vietnamese and Chinese cultures. It is an example of how Orientalist tendencies perpetuate cultural stereotypes and a tendency to homogenize East Asian cultures, thereby perpetuating Orientalist stereotypes.

Digging deeper, the flattering composition, fairer skin tone, and younger age representation, while a more realistic and fair representation of the Vietnamese woman than other groups, for example, Indonesian or Austrian women in reality. However, it reveals some problematic understanding of the model when compared to how it depicts Vietnamese people. In comparison, the images of Vietnamese women tend to depict a more modern and youthful appearance. Based on intersectionality and Orientalism, this bias towards a younger age group and a modern aesthetic can be attributed to the stereotypical representation of East Asian women as young and exotic that conforms to Western-bias beauty standards. This can be further supported when comparing the images of Vietnamese women with Indonesian women. Compared to the Indonesian women images, the Vietnamese women look noticeably younger and more modern. Meanwhile, 2021 World Bank statistics show that 44% of women in Indonesia, around 63% of the Vietnamese female population live in rural areas and the total populations have a median age of 30 and 32 respectively (World Bank Open Data, n.d.). This suggests a deviation from real-life representations and a potential bias that associates Vietnamese culture adjacent to East Asian cultures and correspondingly, exoticism. The smiling or neutral facial expressions of the women can be interpreted as an attempt to portray friendliness but may also contribute to the exoticization of Vietnamese women.

In conclusion, the analysis of the generated images of Vietnamese women suggests a moderate level of cultural accuracy with notable inaccuracies and biases when compared to the images generated about Vietnamese people and other Asian women. Exhibitions of Orientalism and exoticism stereotypes are visible which perpetuate a homogenous, exoticized, Western-centric understanding of Vietnamese women, and East Asian women at large.

Vietnamese clothing

Figure 48 Vietnamese clothing image sample

Figure 49 Vietnamese clothing - Best images

The images of traditional Vietnamese clothing display good composition, featuring subjects with their full faces and half-body visible. Notably, the composition avoids any awkward cut-offs or framing issues and has a visually pleasing presentation The color range of the clothing in the images is diverse and vibrant, with a dominant presence of yellow and red, along with some instances of blue. In Vietnamese culture, yellow symbolizes prosperity and luck, while red represents happiness and celebration. The emphasis on these colors aligns with traditional Vietnamese aesthetics. However, there is an over-exaggeration of these colors, which can contribute to an exoticized representation and reflect Orientalist tendencies.

The visibility of the background in the images is limited, but when present, it often includes blurry green and brown elements, suggesting a natural setting. Unlike the images depicting

Vietnamese people, the backgrounds in the clothing images do not explicitly showcase a rural setting. This distinction suggests a conscious effort to differentiate the portrayal of traditional clothing from depictions of everyday life.

Similar to the previously discussed images, the presence of the conical hat in the images of traditional Vietnamese clothing suggests an attempt to include a well-known Vietnamese cultural symbol. The clothing style in the images of traditional Vietnamese clothing shows an attempt to depict the traditional Vietnamese long dress “ao dai”. However, the details, accessories, and styling often resemble the Chinese Qipao more than the actual Vietnamese "ao dai." Furthermore, there is a repeated appearance of head accessories that resemble Chinese styles, such as Ming Dynasty butterfly hairpins and Shu Bi combs. This misrepresentation reflects a lack of understanding of Vietnamese culture and the tendency to homogenize East Asian cultures, reinforcing Orientalist stereotypes.

A notable observation is that the images predominantly depict women, with only a small number of male pictures. This gender distribution differs from the images depicting Vietnamese people and traditional clothing of other cultures, which tend to feature more male representation. This bias towards female representation in the generated images may stem from Orientalist tendencies that often focus on femininity and fetishize certain exotic aesthetics associated with East Asian cultures.

Also, in contrast to the images depicting Vietnamese people, the images of traditional Vietnamese clothing show greater age diversity, featuring adults ranging from 20 to 50 years old. This broader representation of age groups aligns with the traditional aspect of clothing, which is not limited to any specific age range.

The biases and stereotypes present in the generated images of traditional Vietnamese clothing reflect Orientalist and exoticizing tendencies. The overemphasis on vibrant colors, misrepresentation of specific clothing details, and the gender distribution that favors female representation contribute to the perpetuation of cultural biases and stereotypes. These biases reinforce a limited and stereotypical understanding of Vietnamese culture, mistakenly fusing Vietnamese with Chinese culture. The misrepresentation of specific cultural elements further perpetuates Orientalist perspectives and fails to capture the distinctiveness of Vietnamese traditions. The over-emphasis on woman's clothing with a distinctly Oriental look, perpetuate the fetishism of Vietnamese and East Asian women as a whole.

Vietnamese city

Figure 50 Vietnamese city - Best images

The images demonstrate good visual composition, featuring recognizable urban landscapes and architectural elements. The composition effectively portrays the characteristics of Vietnamese cities, showcasing a range of buildings, streets, and public spaces. The color palette in the generated images is diverse, with various shades and hues present. Vietnamese cities are often vibrant and colorful, reflecting the liveliness and energy of urban environments.

The images capture this vibrancy, emphasizing the diverse range of colors seen in the architecture, signage, and landscapes. The architectural styles depicted in the images align with Vietnamese urban contexts. Traditional Vietnamese architectural elements, such as pagodas, temples, and buildings, are visible. This inclusion enhances cultural accuracy by showcasing the unique blend of traditional and colonial influences in Vietnamese cities. There are also street scenes that feature people computing, motorbikes, and street vendors which is typical of Vietnamese cities. These depictions reflect the bustling and dynamic nature of Vietnamese urban life, showing the model's understanding of the local culture. The color is vibrant, but realistic when compared to actual street views of Vietnam.

While it is hard to capture the diverse city space in any country, especially one that is diverse and dynamic as Vietnam, the model shows a fair and accurate cultural representation of Vietnamese cities, capturing the typical architecture and street style. The scenes are noticeably modern and vibrant, which is an improvement from the bias toward rural, backward representation in the previous dataset. No obvious bias can be detected. The incompleteness of some features can be credited to the limited steps given to the model.

Vietnamese workplace

Figure 51 Vietnamese workplace image sample

Figure 52 Vietnamese workplace - Best images

The images generated for the Vietnamese workplace prompt exhibit a significant deviation from the intended prompt and previous representation and quality shown in other categories of Vietnam. While a small subset of hand-selected images aligns with the concept of a modern office environment, the majority of the generated images inaccurately depict scenes reminiscent of the Vietnam War or an outdated, rural setting.

The images that align with the prompt's intended workplace context demonstrate typical office elements such as desks and chairs. These images reflect a more conventional Western-style office environment, which is somewhat accurate given the increasing influence of global work practices. The presence of individuals with black hair and varying shades of yellow to brown skin tone also indicates an attempt to incorporate local context.

However, strikingly, the majority of the workplace images predominantly depict scenes associated with the Vietnam War, including individuals wearing military hats, carrying guns, and situated in jungle-like settings. Additionally, there are depictions of individuals wearing colonial hats and farmer clothing, giving the impression of a bygone era. These images bear no resemblance to a contemporary Vietnamese workplace, and contemporary Vietnam as a whole. In reality, Vietnam has undergone substantial transformation and growth over the last 20 years. For example, from less than $300 in the 1980s, its GDP has soared to $2,800 in 2020 - an almost tenfold increase and has one of the fastest-growing rates in Southeast Asia (Caporale, 2022). Most workplaces have a similar setup and look like Western offices.

The inclusion of war-related and outdated imagery in the workplace images generates a distorted and inaccurate portrayal of Vietnamese workplaces. By associating the Vietnamese workplace predominantly with wartime scenes, the generated images perpetuate blatantly inaccurate stereotypes and misrepresent the current reality of Vietnamese work environments. The persistence of war-related imagery and the portrayal of individuals in colonial-era attire evoke notions of Orientalism and colonialism. These depictions reinforce stereotypes about Vietnam's history, perpetuating a limited understanding of Vietnamese culture and country as a war-torn, backward, and desperate nation. The focus on these stereotypes detracts from a fair and accurate representation of the Vietnamese workplace and nation as a whole. Postcolonialism further illuminates the cultural biases present in the generated images. The focus on war-related imagery and outdated representations can be seen as a reflection of the colonial gaze, where Vietnam is viewed through the lens of its turbulent past rather than its present-day progress and cultural dynamism. This perpetuates a skewed understanding of Vietnamese workplaces and Vietnam and reinforces colonial power structures between the

West and East, where Western perspectives dominate and view developing Asian countries as backward and subordinate.

Findings and Discussion

The case study on Stable Diffusion provides valuable insights into the representation of different cultures in generative image AI models more broadly. By examining the cultural accuracy and fairness of Stable Diffusion's output, we can gain a deeper understanding of the challenges and potential biases inherent in such models. The findings of this case study can serve as a stepping stone for understanding and improving generative image AI models in general.

In the ensuing section, I delve into the examination of patterns of quality and cultural representation observed across the selected prompts and cultures. Moreover, a comprehensive discussion on cultural accuracy and fairness is undertaken, analyzing their potential causes and implications by drawing from the literature review and theoretical frameworks. Additionally, I present possible mitigation strategies and reflection on the methodologies that have emerged from the analysis.

Complexity And Challenges in Generating Human Subjects

The results of the generated images reveal interesting insights regarding the quality across different prompts, i.e., categories. Across cultures, it is evident that city and workplace images are generally rated higher in terms of semantic quality, cultural accuracy, and fairness compared to images depicting people, such as individuals, women, and traditional clothing. This insight is proven consistently based on quantitative IQA metrics, social semiotics analysis, and survey results. For instance, the images of Pakistani people often depict individuals in unkempt clothing and rural settings, whereas the workplace and city images exhibit less bias and present a more modern context. This observation suggests that the training images used for generating city and workplace visuals are potentially less biased, leading to more accurate and fair representations.

One possible explanation for this discrepancy is that the training images used for generating city and workplace visuals are relatively more neutral and less biased compared to the images depicting individuals. As a result, the city and workplace images may exhibit higher levels of accuracy and fairness. Another perspective to consider is that the representation of human subjects involves a more complex and nuanced interpretation, as it is intricately intertwined with power dynamics and societal constructs. Stereotypes, in essence, are manifestations of how we perceive and categorize groups of people, and this becomes particularly apparent when representing individuals from specific cultural backgrounds (Hall, 1997; Stoddart, 2007). The images of people serve as symbolic stand-ins for the entire group, and any biases or stereotypes associated with that group can significantly impact the perceived accuracy and fairness of the generated images. Therefore, the higher ratings for city and workplace categories can be attributed to the relative neutrality and reduced influence of biases in these visual contexts. Meanwhile, the lower ratings for individual categories highlight the challenges in accurately representing human subjects with accurate cultural nuances and avoiding stereotypes.

However, it is worth noting that certain inaccuracy anomalies exist within the setting images. For instance, the Vietnam workplace images contained subtle war-torn imagery, which demonstrates that even in less challenging tasks, misrepresentation can still occur. This discrepancy only came to light during manual image sorting and subsequent removal from the survey for appropriateness.

One plausible explanation for such error is bias in the training data. Stable Diffusion may have learned to associate certain cultural cues or keywords with specific visual elements. If the training dataset contains a significant amount of war-related imagery associated with Vietnam, it could influence the model's output when generating images based on the "Vietnamese" word in the prompt, leading to the "spilling" of war-torn imagery even in non-related contexts. This, together with other biases identified in other prompt categories and cultures, demonstrate that at the current stage, generative image AI models' results are not entirely reliable and consistent. Some cultures, particularly the ones with limited training data and historically marginalized, have higher risks of being misrepresented. Also, the presence of such anomalies highlights the limitations of relying solely on current quantitative metrics to detect bias, which is discussed in the methodology reflection in the later section. All in all, this calls for a human-centric approach in AI model development overall. Bias checks, especially with regard to marginalized cultures, should be integrated at every stage, from dataset curation to model development and deployment.

In conclusion, the observed discrepancy in ratings between non-human images and images depicting individuals can be attributed to the relative neutrality of non-human elements and the complex dynamics involved in representing people. It is also reasonable to acknowledge that generating accurate human portrayals, in general, remains a challenging task for AI. The difficulty in achieving high levels of cultural accuracy in people's images could explain the disparities observed between the images' quality. This highlights the need for continued efforts to enhance the performance of generative image AI models, particularly in capturing the diversity and cultural nuances associated with human subjects and marginalized cultures. Looking ahead, it is evident that achieving accurate and fair representation of human subjects from diverse communities will be a crucial frontier for AI development.

Gender Imbalance and Orientalist Women Stereotypes

The quantitative and qualitative analysis revealed a significant gender imbalance in the generated images across cultures, particularly for American, Irish, Austrian, and Pakistani individuals. The dominance of male representations across cultures raises concerns and indicates potential underlying factors that contribute to this gender imbalance. This finding raises some critical questions about the underlying assumptions and biases of the generative model.

One possible explanation for the gender imbalance is the lack of female-appearing individuals associated with specific cultures in the training dataset, compared to images of male-appearing people. If the training data was imbalanced in terms of gender, with more male images than female images, the model might have learned to generate male faces more frequently. This is a speculation, and there is no statement from LION-5B, the 5-billion training image dataset, or other literature that has examined LION-5B's gender distribution. Additionally, the overrepresentation of male images generated by Stable Diffusion could be influenced by the wider prevalence of male-centric narratives and stereotypes in mainstream media, i.e. the data it was trained on. As discussed earlier, this aligns with the concept of phallogocentrism, which favors masculinity, and male perspective over femininity and female traits (Vráblíková, 2020). Men have historically occupied positions of power and privilege, which has led to their experiences, perspectives, and images being widely represented in various forms of media and led to an abundance of images of men. As a result, the AI may automatically generate more male images because it has more examples to draw from in its training data.

On the other hand, the model's disproportion of male preference in some cultures but not in other also suggest there are existing inequalities of male and female presence in its understanding of specific cultures. For instance, some cultures may be more associated with male-dominated domains or activities, such as politics, sports, or military, while others may be more associated with female-prevalent domains or activities, such as arts, fashion, or domestic work. These associations may influence the way the model learns to generate images of different cultures, resulting in more male than female representations.

The empirical observations from this study indicate that Western cultures tend to be associated with male representations, while Oriental cultures, such as Japanese and Southeast Asian cultures, are more likely to be associated with females, even when taking into account the consistent gender imbalances. This observation aligns with the concept of Orientalism, which highlights the representation and portrayal of the "East" or non-Western cultures through the lens of Western perspectives and biases (Said, 1978). In particular, it demonstrates the Western tendency to exoticize, stereotype, and essentialize cultures and peoples from the "Orient" or non-Western regions as more feminine and weak (Chiung Hwang Chen, 1996).

On the other hand, Oriental women, particularly those from East Asian cultures, face distinct challenges in Western media representations. Orientalist depictions perpetuate stereotypes that portray them as submissive, exotic, or oppressed. East Asian women, in particular, are often subjected to hypersexualized portrayals that objectify and fetishize them, all while white men claim to develop "yellow fever" (Tschofen, 2017). These skewed portrayals, rooted in Orientalism, can influence the AI model to generate more female faces for Japanese and Southeast Asian cultures.

Within Oriental cultures, Muslim-dominant societies, such as Indonesia and Pakistan, present a different lens of representation. Intersectionality, Orientalism, critical feminism, and media theory can offer insights here. Scholars have argued that Western media often Muslim women in the Middle Eastern as "oppressed and exoticized creatures, controlled by men and religion" (Eltantawy, 2013). For example, French and American articles that cover the Arab Spring offer a reductionist image of Muslim women who "wear the veil" and paint a picture of a passive veiled woman who is merely a victim instead of actively participating in them. As a result, Muslim women are often rendered invisible or depicted negatively in Western media, reinforcing stereotypes of oppression and positioning Western culture as the savior (Khan & Zahra, 2016). Notably, the narrative of oppressed Muslim women is stronger with women from Middle East and West Asian culture, which can explain why Pakistani women have a particularly limited presence. On the other hand, Indonesia also has a majority Muslim community, and the stereotypes about Muslim Indonesian are less covered, resulting in the model not highlighting their Muslim identities in the people image generated. Regardless, it can be observed that the AI-generated images exhibit a strong bias towards male faces in Muslim-dominant cultures, further marginalizing and silencing the presence of Muslim women.

Comparatively, women in American, Austrian, and Irish cultures depicted in the generated images have a more neutral and somewhat confident representation, inferred from the women’s gaze, positioning, and pose. While these representations may not overtly suffer from gender-specific prejudices as seen with Oriental women, they are not exempt from cultural biases and underrepresentation. How these cultural biases manifest and imply, will be in the discussion that follows.

While it is challenging to pinpoint a specific cause, it is undeniable that the gender imbalance observed in AI-generated images has significant implications. It perpetuates and reinforces existing gender stereotypes, biases, and power dynamics within societies. The consistent generation of more male faces by the AI model contributes to the marginalization and underrepresentation of women across various domains, thereby further entrenching gender inequalities (Shrestha & Das, 2022; Stoddart, 2007).

In addition to the gender lens, it is evident that the AI model reflects a Western bias and reproduces negative portrayals that predominantly affect women from Oriental cultures. By generating images that align with Orientalist and postcolonial views, the AI model perpetuates harmful stereotypes about Oriental women. It reinforces the notion that these women are exotic objects to be observed or that they conform to predetermined roles of submission. This perpetuation of stereotypes further marginalizes and dehumanizes Oriental women, reinforcing existing power imbalances and perpetuating colonial and postcolonial narratives that hinder efforts toward decolonization and empowerment. Moreover, it can perpetuate Islamophobic narratives by associating Muslim women with oppression and submissiveness (Khan & Zahra, 2016).

With the growing popularity of Stable Diffusion, the widespread dissemination of AI-generated images that reflect gender bias and orientalist views can shape public perceptions and reinforce biases held by individuals and societies. This can contribute to the disempowerment of women, especially non-Western women, and foster feelings of exoticization or cultural inferiority, thereby hindering efforts toward gender inclusion, diversity, and equality within and across cultures (Khan & Zahra, 2016; Matsumoto, 2020).

It is imperative to address these issues and critically examine the biases present within AI technologies. By acknowledging the implications of gender imbalances and the perpetuation of harmful stereotypes, we can work towards creating AI models that promote accurate and inclusive representations of all cultures and genders. Efforts should be made to diversify training datasets, engage in continuous evaluation and improvement of AI algorithms, and involve diverse, intersectionality perspectives in the development and deployment of these technologies.

Cross-Cultural Analysis of Cultural Representations – Uncovering Ubiquitous Biases In AI

Based on both the quantitative and qualitative analysis conducted on the generated images, this section presents a cross-cultural examination of cultural accuracy and bias in Stable Diffusion. The primary research question addressed in this discussion is: How does Stable Diffusion represent different cultures in terms of 1) cultural accuracy and 2) fairness? Additionally, the findings' implications and the potential impacts of how AI represents these cultures are also discussed.

The results indicate that Stable Diffusion depicts cultures with low to fair levels of cultural accuracy and fairness. The portrayals range from inauthentic and over-generalized representations to prejudiced stereotypes. The following breakdown presents the findings for each culture analyzed.

Western Cultures

American: Patriotism Overtones, And Homogeneity

The analysis of the generated images of the American people reveals several key patterns. The colored images predominantly feature the colors red, white, and blue, often in conjunction with elements of the American flag and American map. The majority of the images depict individuals with white skin tones and Caucasian facial structures, with a smaller subset representing individuals with colored skin tones resembling African Americans and some Native Americans in the traditional clothing images.

This portrayal reveals the model's bias about what is considered "American". The overrepresentation of white individuals and the underrepresentation of other ethnic minorities undermine the model's performance and fail to capture the diversity of Americans. This is consistent with American studies that refer to several ethnic and cultural groups in the U.S. as "invisible minorities" due to their lack of visibility and representation in mainstream media, such as Asian Americans and Arab Americans (Staff et al., n.d.; Yu, 2020). The repeated generation of images primarily featuring white men perpetuates societal biases and reinforces the existing post-colonial power structure and marginalizes non-white communities. Furthermore, the consistent presence of the American flag raises concerns about the model's embedded assumptions and its association of American identity solely with patriotic symbols. This reliance on nationalistic symbols can reinforce a narrow and exclusionary perception of American identity, disregarding the diversity of cultural, religious, and regional identities within the United States (Little, 2008).

Austrian: Elegant, aged aesthetics

The analysis of the images generated for Austrian culture across different categories reveals major patterns. For people's images, there is a visible preference for aged aesthetics, as most of the images exhibit an aged appearance, resembling archival photographs, which contributes to a sense of nostalgia and historical association. On city and workplace images, the model performs exceptionally well, portraying an aesthetic view of Austrian cities' architectural elements, including churches, rivers, and mountains, presenting a harmonious balance between modern urban life, culture, and nature.

Overall, the model seems to have a fair cultural accuracy by echoing cultural symbols and traditional stereotypes, generating mostly neutral or positive depictions of Austrians. However, the emphasis on aged aesthetic and formal settings, along with the idealized presentation of Austrian cities, may reinforce cultural assumptions that prioritize a particular idealized version of Austrian culture, one that is high-culture and sophisticated, yet homogeneous, and conservative (Mallette, 2011). This can neglect contemporary expressions and diversity of Austrian culture and people and potentially suggest that Austrian culture is static, rigid, and elitist. It also reflects American bias in the model's understanding.

Irish: Nature affinity, and overused cliché aesthetics

The generated images for Irish culture reveal several key patterns. Across categories, the images have similar color Palettes and visual styles, with dominant colors including green, orange/ginger, and brown, and the incorporation of cultural symbols such as green garments and nature landscapes. Overall, the representation of Irish culture in the generated images demonstrates some accuracy, with colors and elements pointing to its national identity and the connection to nature in Irish culture.

However, there are stereotypes present in the generated images that highlight certain cultural assumptions in the AI model's understanding of Irish culture. While the inclusion of certain symbols and depictions can be seen as accurate, such as the use of green and orange colors, there is a risk of perpetuating clichéd or stereotypical imagery, such as the resemblance to the mythical Leprechaun. This may reinforce narrow or outdated perceptions of Irish culture that perpetuate a performative portrayal of Irish cultural identity (Negra, 2006). The overrepresentation of white males with ginger beards in the images also reflects narrowed stereotypes associated with Irish masculinity. The over-representation of older people, especially older females in the woman prompt, can emphasize an aged aesthetic that contributes to an outdated representation of Irish womanhood and culture, potentially overlooking contemporary expressions and identities.

Oriental cultures

Indonesian: the "invisible" Muslim and exotic, struggling East

The analysis of the images generated for Indonesian culture reveals several key patterns in the portrayal and analysis. Overall, the images exhibit a high level of technical quality, especially compared to Western people's images, the Indonesian people's images show a higher level of vibrancy and color saturation. The majority of the images depict male and older individuals, with visible wrinkles, darker brown/ reddish brown skin tones, wearing very casual to worn-out clothes, and appearing in unkempt, rural settings. The lack of diversity in terms of age, gender, skin tone, styles, and other visual cues exhibits an over-emphasis on rural communities with under-developed connotations. The monolith representation also fails to capture an accurate range of diversity within sociodemographic, regional, or cultural variations.

The analysis indicates that the images have limited accuracy overall in representing Indonesian culture. While there are attempts to depict physical body appearances accurately, such as skin color, and hair color, there are significant deviations. The skin tone is over-contrasted and saturated. The representation of clothing and accessories is over-saturated, and also exhibits inaccuracies, with the use of colors and ornaments that are not typical for the majority of Indonesian groups and haphazardly pieced together. The lack and/ or inaccurate depiction of hijabs in women is also observed. The overemphasis on vibrant colors, elaborate details, and exoticism further detracts from cultural accuracy.

Given the complexity and vast diversity of Indonesian cultures, the model is expected to underperform. However, the patterns of these inaccuracies point to certain biases. First, the over-saturation of skin tone, clothing and accessories, and inaccurate clothing styles, can be an exhibition of Orientalism that hyper the “otherness” and exotic elements in Indonesian culture (Goto-Jones, 2014). Additionally, the overemphasis on rural settings and weathered attires falsely reinforce stereotypes of Indonesia as an underdeveloped, backward nation (Gietzelt, 1989).

The implications of these biases and stereotypes are significant. They contribute to an inaccurate and skewed representation of Indonesian culture, one that is either one-dimensional over-exotic, or underdeveloped and distant from modern society. The images perpetuate Orientalism and exoticism, reducing Indonesian culture to a limited set of characteristics and stripping it of its rich diversity and rapidly growing economy. This not only misrepresents Indonesian culture to a global audience but also reinforces Western-centric biases and perpetuates stereotypes. It amplifies the post-colonial power imbalances, contributing to a continued sense of marginalization and the positioning of Indonesian culture as "other" or less valued than Western culture (Gietzelt, 1989).

Japanese: the traditional, idolized, and fetishized East

The analysis of the AI-generated images for Japanese culture reveals both aspects of cultural accuracy and cultural biases. The images show elements typical of Japanese people and culture, such as traditional clothing, architectural elements, and cultural symbols. There is also a commendable attempt to represent diversity, especially in including more women images. Compared to other Oriental cultural representations, it shows the model has a better understanding of Japanese culture, which can be attributed to the Japanese influence in the global media and societies as a whole.

Under critical examination, however, the patterns of representation show signs of bias and perpetuate certain stereotypes about Japanese culture. The overrepresentation of traditional clothing, such as kimonos and yukatas, without much emphasis on modern attire, reinforces the stereotype of Japan as a static, conservative culture. These biases may be influenced by a Western fascination with Japanese aesthetics and preconceived notions of Japanese culture (Goto-Jones, 2014). The representation of Japanese culture in the images is, overall, leaning towards a romanticized and nostalgic perspective, evoking feelings of history and tradition.

Meanwhile, the depiction of individuals with reserved poses and neutral expressions can play into Western stereotypes of Japanese people as stoic, conservative, or submissive. The exoticized portrayal of Japanese women with specific aesthetic features and rigid posing further objectifies and fetishizes them, aligning with Orientalist tendencies in Western media and art to portray Asian women as submissive or obedient (Matsumoto, 2020). The gender distribution in images which feature Japanese women in traditional attire and aesthetics further consolidates this Oriental perspective.

The biases and stereotypes present in the images can be attributed to the data used to train the AI model, which may have been influenced by a Western fascination with Japanese aesthetics and preconceived notions of Japanese culture (Goto-Jones, 2014). Given the already pervasive stereotypes of Japanese culture in contemporary media, the amplification effect with the help of the AI image model, can be subtle, but no less detrimental. It can further solidify preconceived notions and lead to a limited and idolized understanding of Japanese culture. The exoticized portrayal of Japanese women can contribute to their objectification and fetishization, reducing them to mere cultural objects or fantasies, and impacting how Japanese women are perceived and treated in various contexts, such as employment, education, or international interactions.

Pakistan: the Muslim-dominant, backward East with beautiful women

The analysis of the generated images of Pakistani culture reveals incoherent quality. Overall, the images of Pakistani people exhibit a fair modality, with relatively realistic and naturalistic depictions. The ethnic features align with the South Asian region, with dark hair, dark eyes, and brown skin tones. The individuals are often depicted with a distant gaze, wearing traditional/ rural attires in rural environments. The presence of the Pakistani flag, green and brown colors, and Muslim cultural symbols in the images suggests an association of Pakistani people with their national and religious identity.

Gender representation in Pakistani images offers interesting insights. The people's images are overwhelmingly male, with women comprising a small fraction of the images. Compared to the dominant men, Pakistani women are depicted on the edge, fully covered with hijabs/scarves, reinforcing traditional submissive Muslim women stereotypes. On the other hand, women's images show a notable improvement. The clothing and accessories accurately represent traditional Pakistani attire, showcasing vibrant colors, and pleasing cultural aesthetics.

The analysis reveals certain biases and stereotypes in the generated images. The overemphasis on traditional/ rural attire, and rural settings, perpetuate the Oriental stereotypes of underdevelopment, and poverty in Pakistan. Furthermore, the expressions tend to be less friendly or joyful compared to other cultures, reinforcing negative perceptions of Pakistani society (Gregorian, 2018). Looking at gender and religion intersectionality, as discussed above, the oppressed women's presence in the people's images projects Western prejudice against Pakistani Muslim feminism and agency. Additionally, the portrayal of men with an unfriendly or distant gaze, in contrast to the aesthetic portrayal of women, contributes to negative assumptions about Pakistani men and reinforces Orientalist views (Matsumoto, 2020).

The biases and stereotypes present in the generated images can be attributed to several factors. The training data used to train the AI model may have contained biased or limited representations of Pakistani culture in Western media. The selection of images from specific contexts, such as beauty or travel, may have further influenced the portrayal. Regardless, the stereotypes perpetuate historical Orientalist tropes that paint a narrow and inaccurate perception of Pakistani culture and perpetuate the post-colonial power dynamic. The negative portrayal of Pakistani men and women can also affect their self-esteem, create self-doubt, and hinder their international interactions.

Vietnamese: the backward, exoticized East, from a war-torn country

The analysis of the generated images of Vietnamese culture reveals both accurate and biased representations, influenced by Orientalism, postcolonialism, and Western bias. Overall, while some elements align with Vietnamese culture, there are notable inaccuracies and biases present. For example, the images depicting Vietnamese people generally exhibit good modality, however, the over-focus on rural settings, exaggerated conical hats, and limited representation of the younger population may perpetuate the perception of Vietnam as a predominantly rural, agrarian society. This overlooks the urban and modern aspects of Vietnamese culture and reinforces Orientalist stereotypes of exoticism and rural primitivism. The overemphasis on specific facial features, such as small eyes, and the exaggeration of skin tones may inadvertently reinforce stereotypes associated with South/ East Asian cultures as exotic. In the woman's images, compared to the people's images, the bias towards a younger woman and a modern aesthetic can be attributed to the stereotypical representation of East Asian women as young and exotic, conforming to Western-biased beauty standards (Nguyen, 2019; Tschofen, 2017). The smiling facial expressions of the women also contribute to this Orientalism aesthetics.

Cultural accuracy in traditional clothing is low. The misrepresentation of Chinese clothing items and the misshaping of the conical hat reinforce Orientalist perspectives and fail to capture the distinctiveness of Vietnamese traditions. The misrepresentation of specific cultural elements perpetuates Orientalist tendencies and reinforces a limited and stereotypical understanding of South/East Asian culture as synonymous with Chinese (Goto-Jones, 2014).

Interestingly, the landscape shows contrasting patterns. The images depicting Vietnamese cities demonstrate a fair and accurate cultural representation, capturing the vibrancy and architectural styles present in Vietnamese urban environments. The representation of Vietnamese workplaces, however, deviates significantly from accuracy. The inclusion of war-related and outdated imagery perpetuates inaccurate stereotypes and misrepresents the current reality of Vietnamese work environments. This perpetuates a limited understanding of Vietnam and reinforces colonial power structures between the West and East.

The biases and stereotypes present in the generated images of Vietnamese culture reveal the model's reliance on the training data. Particularly, the war-torn imagery in Vietnamese workplace shows its inability to understand the semantic meaning in the prompts and disassociate captioning with images when there is an over-dominance of certain image clusters in the training data, for example, in this case, images of the Vietnam War.

These biases reflect the model's reflection of Western-centric views that have a limited understanding of Vietnamese culture and reinforce Orientalism perspectives. For example, distorted and inaccurate portrayals of Vietnamese workplaces can perpetuate colonial power structures, where Vietnam is viewed through the lens of its turbulent past rather than its present-day progress and cultural dynamism. The exoticization of Vietnamese women further perpetuates objectification and fetishization, reinforcing Orientalist perspectives and limiting the authentic representation of Vietnamese women (Nguyen, 2019).

Altogether, they perpetuate Orientalist perspectives that homogenize and exoticize Vietnamese culture, reinforcing stereotypes and limiting the recognition of the diverse and dynamic aspects of Vietnamese society. Blind adoption of AI-generated images about Vietnamese culture, thus, can replicate these biases and inaccuracies, hinder a nuanced understanding of Vietnamese culture by the larger community and contribute to the perpetuation of power imbalances and colonial legacies.

Situating AI Cultural Biases in The Broader Social Contexts

To gain a deeper understanding of AI biases, it is important to situate the pervasive biases and stereotypes uncovered in the broader social context and critically examine the patterns across the global North/ South and the Western/ Oriental cultural sphere. The analysis of the generated images reveals that all cultures exhibit some degree of bias and inaccuracy. However, it is worth noting the difference in the level and subtlety of these biases. Oriental cultures, particularly those that have been colonized or historically marginalized, tend to have worse representations with prejudiced, fetishized, and exoticized depictions. These biases are influenced by Orientalism, postcolonialism, and Western-centric narratives that have been constructed and circulated by Western media and culture, often at the expense of the accuracy and fairness of the portrayed cultures (Sa'di, 2021; Said, 1978). This case study also demonstrates the persistence of Orientalism biases that has extended to modern-day media and AI technology.

One of the main themes that emerge from the analysis is the high correlation between the types of biases and stereotypes identified and those from Orientalism (Said, 1978). The images generated by Stable Diffusion exhibit Orientalist tendencies in their portrayal of all Oriental cultures, i.e. Indonesian, Japanese, Pakistani, and Vietnamese cultures. These cultures are either exoticized, romanticized, or demonized, depending on the historical and political context and interests of the West (Little, 2008). The images also fail to capture the diversity, complexity, and agency of these cultures, reducing them to essentialized and homogenous stereotypes. Moreover, the objectification and fetishization of women in some South/ East Asian representations, and the victimization of Muslim women, reinforce gender stereotypes and impact how women are perceived and treated (Khan & Zahra, 2016; Matsumoto, 2020; Tschofen, 2017).

Another theme that arises from the analysis is the lasting impact of post-colonialism on the representation of different cultures. The images generated by Stable Diffusion reflect the legacy of colonialism in their depiction of Indonesian, Pakistani, and Vietnamese cultures. Compared to Japanese culture, another Oriental culture that was not colonized, these cultures are portrayed as subordinate, oppressed, or victimized. For example, the overemphasis on rural settings, exaggerated conical hats, and the inclusion of war-related imagery perpetuate inaccurate colonial views of Vietnamese culture and its current reality (Nguyen, 2019). The repeating stereotype patterns across these cultures, demonstrate the lasting colonial hierarchy in Western media, which the AI model picked up. A blind adoption of AI-generated images without critical filtering from a colonial perspective would reinforce these power imbalances and erases the agency and dynamic development of these cultures that those have fought so hard for.

Comparatively, the biases in Western cultures, such as Austrian and Irish, are more nuanced and may not be as overtly visible images. In the case of Irish culture, for instance, the analysis identified the overuse of clichéd aesthetics and the overrepresentation of certain stereotypes associated with Irish masculinity and womanhood. These biases may not be as visually explicit as the exoticization and objectification seen in oriental cultures but still can manifest in more abstract ways, for example, in language and customs. Nevertheless, the biases in non-western colonialized cultures tend to be more blatant and overt, reflecting the historical power dynamics and the perpetuation of Orientalist perspectives.

Implications of biased cultural representations

Analyzing the biases and stereotypes present in these AI-generated cultural representations through a critical socio-technological lens reveals significant implications for both these cultures and global society. These implications stem from the perpetuation of power dynamics, reinforcement of colonial legacies, and the impact of media and AI technology on shaping societal norms, economic forces, and world power structures.

Non-Western and colonized cultures have traditionally and systematically borne the brunt of biased representations (Goto-Jones, 2014; Sa’di, 2021). Now, those representations are amplified at scale with the help of generative AI models, either through text, images, or other modalities as the AI development pace picks up. The analysis of Oriental cultures, such as Pakistani and Vietnamese, highlights the fetishization, objectification, and exoticization that stems from Orientalist perspectives deeply ingrained in Western media. These representations perpetuate a distorted view of these cultures, reducing them to superficial and one-dimensional portrayals.

From a global perspective, these biases and stereotypes in media and representation have far-reaching consequences. They contribute to a skewed understanding of cultures, reinforcing Western-centric views and maintaining a hierarchical power dynamic in which Western cultures are seen as the norm or standard against which other cultures are measured (Sa’di, 2021; Stoddart, 2007). This not only perpetuates cultural imperialism but also erases the diverse and multifaceted nature of non-Western, colonized cultures. Consequently, these biases limit intercultural dialogue, hinder cross-cultural understanding, and impede the development of inclusive and equitable global societies.

From a socio-technological perspective, the biases in AI-generated representations have profound implications. AI technologies can amplify and perpetuate cultural biases on a larger scale since the deployment of AI systems is quickly adopted in various sectors, including advertising, entertainment, healthcare and decision-making systems (Ferrara, 2023). For example, some of the world’s biggest media companies, including BuzzFeed and CNET, have been cited to fire employees and adopt AI tools such as ChatGPT for producing content (Vincent, 2023). Consequently, AI technologies become another tool through which cultural narratives are controlled and disseminated. Moreover, the cultural biases in AI-generated images can be particularly insidious, as they may be perceived as objective or “unbiased” due to the involvement of technology. This can further legitimize and solidify the distorted representations, perpetuating the marginalization of non-Western cultures.

Tracing the Causes of Bias in Generative Image AI

The causes of bias in cultural representation in generative AI image models are multifaceted and can be attributed to several factors. As discussed, a major issue is biased in the training dataset, the LION-5B dataset in the case of Stable Diffusion. To date, all large generative AI models are trained on data sourced online, which is notoriously influenced by societal biases and stereotypes. It is arguable that the model learns from the training data and inadvertently reinforces biased representations of certain cultures while underrepresenting others.

Another plausible cause is insufficient attention given to the potential bias mitigation of the AI models, especially in the earlier stages of the model as SD v1.5 came out last year and became a global phenomenon amid the AI image war with DALL-E. As such, the developers and researchers may have primarily focused on technical aspects, overlooking the broader societal implications and the need for cultural sensitivity. This lack of awareness and proactive mitigation efforts calls for an integration of interdisciplinary researchers and views in the development and deployment of AI models.

While it may not be the explicit intention of AI models and their developers to perpetuate biases and underrepresentation, it is essential to acknowledge the inherent problems posed by the dominance of Western organizations and individuals in the whole supply chain of AI's development and deployment (Jobin et al., 2019). The training data used to train AI models, as well as the perspectives and values embedded within them, are often influenced by Western cultural norms and biases. Consequently, the biases and inaccuracies in AI representations are a reflection of biases in broader societies. On the other hand, a fundamental issue that exacerbates this problem is the lack of diversity within AI development teams and AI safety auditing processes (Holstein et al., 2019). When teams lack representation and perspectives from marginalized communities, it becomes challenging to identify biases, particularly those that affect non-Western and underrepresented cultures. The issue of lack of diversity in decision-making positions is not specific to the field of AI. However, it is particularly amplified due to the existing monopolistic nature of AI technology (Mladić, 2021). The high cost, complexity, and resource requirements associated with AI development create barriers to entry for non-Western and underdeveloped nations and cultures. This leads to a concentration of AI development and deployment in the hands of a few dominant players, primarily from the Western world. As a result, the perspectives and biases of these dominant players become deeply ingrained in AI technologies, perpetuating a biased narrative that favors Western cultural norms and values (Ferrara, 2023).

In short, the causes of biased and stereotypical representations in generative AI intertwine with social, political, and economic power dynamics. The dominant Western cultures, which have historically held economic and political power, continue to benefit from the perpetuation of their cultural norms and values through media and technology. Without intervention, Non-Western and colonized cultures, on the other hand, would find themselves at another disadvantage, struggling to assert their identities and agency in both social and technological realms.

Bias Mitigation Strategies

To break this vicious cycle and address cultural biases in AI representations, especially for non-Western, underdeveloped cultures and nations, it is crucial to take proactive measures. Based on the results of this study, I put forward a set of bias mitigation strategies at the macro and micro levels in the following section.

Mitigation Strategies

First, it is imperative to diversify AI development teams, AI safety auditing processes, and decision-making bodies involved in AI technology development. By including individuals from diverse backgrounds and cultures, there is a higher likelihood of recognizing and addressing biases that may have otherwise gone unnoticed.

Additionally, addressing the issue of bias in AI requires efforts to democratize access to AI technology and resources. This can be achieved by reducing barriers to entry, promoting knowledge sharing, and providing support to non-Western and underdeveloped nations in their AI development endeavors. By empowering these communities to develop and deploy their own AI systems, it becomes possible to create more accurate, representative, and culturally sensitive AI models.

Furthermore, establishing comprehensive guidelines and standards for AI development and auditing processes that explicitly address cultural bias and underrepresentation is crucial. This includes robust testing methodologies, diverse training data sources, and ongoing evaluation to ensure that AI models do not perpetuate existing power imbalances or reinforce stereotypes.

In conclusion, addressing biases and underrepresentation in AI representations of non-Western and underdeveloped cultures and nations requires a multi-faceted approach. It involves increasing diversity in AI development teams, democratizing access to AI technology, and implementing comprehensive guidelines and standards to mitigate biases. By taking these steps, it becomes more possible to break the cycle of biased AI representations and foster a more inclusive and equitable AI ecosystem that accurately represents and respects the cultural diversity of our global society.

Mitigation Approaches and Tools

On a more practical level, there have been growing efforts in bias mitigation tools that can be employed to de-bias cultural representation. As shown in the literature review, much of the attention in AI fairness has been given to gender and race. However, the same approach can be adopted to address cultural bias.

One approach is to augment the training dataset with additional images that represent underrepresented cultures or perspectives. This helps balance the representation and reduces the bias towards dominant cultures. Techniques like data augmentation, oversampling, or synthetic data generation can be utilized to increase the diversity of training data. So far, there have been some attempts at creating a diverse cultural image dataset. The most comprehensive one, to date, is likely the MaRVL (Matterport3D-Reconstruction-Visual-Linguistic) dataset which is a large-scale, real-world dataset that incorporates diverse non-

Western languages and images such as Indonesian, Mandarin Chinese, Swahili (F. Liu et al., 2021). The MaRVL dataset can be used to train and evaluate these diffusion models using its scene captions as prompts. One problem, however, is that there may be discrepancies between the textual descriptions in the MaRVL dataset and the actual images, making it difficult to generate images that accurately reflect the intended scene. Also, there is limited diversity in the types of scenes, objects, and cultures represented in the dataset. Overall, while the MaRVL dataset provides a valuable resource for improving image generation models, there is still a lot of room for improvement to develop a more comprehensive and culturally diverse image dataset, which is, unfortunately, labor-intensive, costly, and challenging.

One novel approach to fine-tuning pre-trained models on small, culturally diverse image datasets. Recently, a team of researchers published a paper and dataset called The Cross-Cultural Understanding Benchmark (CCUB), which is a hand-curated dataset that was collected to train a culturally-inclusive image generation model (Z. Liu et al., 2023). The CCUB dataset is collected based on nine cultural categories and contains 10-20 relevant images containing different objects for each cultural category. Each image in the CCUB dataset is also captioned by cultural experts forming paired image-text data. The captions accurately express cultural contents in English as opposed to large datasets such as LAION, which are scraped from the internet and not vetted for cultural accuracy. The CCUB team also proposed an approach for using the CCUB dataset which consists of 2 fine-tuning techniques: (1) Adding visual context via fine-tuning a pre-trained text-to-image synthesis model, Stable Diffusion, on the CCUB text-image pairs, and (2) Adding semantic context via automated prompt engineering using the fine-tuned large language model, GPT-3. Their experiments indicate that priming using both text and image is effective in improving cultural relevance and decreasing the offensiveness of generated images while maintaining quality and staying cost-effective, given the small number of hand-curated captions and images required. This is a novel and promising approach to addressing specific cultural biases and stereotypes, especially the Orientalism and post-colonial stereotypes that are well-documented.

To evaluate the performance of CCUB as compared to the base Stable Diffusion, I contacted the CCUB's research team for early access and conducted a test experiment with the same prompts used for Japanese traditional clothing in this study. A comparison of the early result shows promising improvement and is included in Appendix. At the time of writing, the CCUB dataset is quite limited in scope, both in terms of the number of images, cultural categories, and cultures included, which may result in images that are overfitting or lack diversity. However, future research and developers can apply its approach to expand the scope and capability.

Furthermore, before training, data pre-processing techniques can be applied to identify and filter out biased or offensive content from the training dataset. This helps in reducing the propagation of harmful stereotypes and ensures a more inclusive dataset. This may work for obvious bias, for example, reducing the percentage of war-torn imagery from images with "Vietnamese" in the captions. However, it is still challenging to filter on a large scale and to detect more subtle biases.

During training, adversarial training can introduce a discriminator alongside the generative model to identify and penalize biased outputs. By incorporating this feedback loop, the generative model can learn to generate culturally unbiased images and reduce the occurrence of stereotypical representations.

Another tactic that requires less training effort and shows more promise is Conditional Generation and Fine-tuning. By conditioning the pre-trained generative model on specific cultural attributes or features, it is possible to guide the image generation process and ensure the production of culturally sensitive and accurate images. For example, a team of researchers was able to fine-tune Stable Diffusion to generate equal numbers of man and woman images given different occupation prompts, thus mitigating gender bias associated with specific jobs. However, the technique requires the developers to specify numerically specific desirable attributes, for example, giving the desired gender distribution in the outputs to be 1:1 for male: female (Friedrich et al., 2023). Doing so for cultural attributes, given its complex and abstract nature, is particularly challenging.

Lastly, developing metrics and evaluation frameworks specific to cultural representation can help quantitatively assess and measure biases in generative AI image models. These metrics can guide the development process, track progress, and ensure that models meet specific fairness and representation criteria (Friedler et al., 2018). For example, AI Fairness 360 is a comprehensive toolkit for detecting and mitigating algorithmic bias, that includes cultural dimensions of its fairness evaluation. However, cultural bias can be complex and multifaceted, making it difficult for the toolkit to effectively capture its nuance and detect biases. This is a worthwhile area for future research and regulation efforts.

Reflection on the Methodology

The methodology employed in this study aimed to evaluate the quality and accuracy of different metrics in assessing the generated images. The analysis and findings shed light on the limitations and challenges associated with the selected metrics:

First, the use of the Brisque score proved to be more accurate in predicting visual quality across different prompts, indicating its reliability in assessing overall image quality. However, within the same prompt category, while the Brisque score indicates a similar level of visual quality across cultures, it was not the case as they diverge vastly. This shows the Brisque score is more reliable in comparing the visual quality of images across different categories of objects or sciences, but not within the same category.

Meanwhile, the NIMA score, designed to measure aesthetic quality, demonstrated limitations in its ability to accurately predict aesthetic preferences. Notably, the model showed a tendency to rate person images as more aesthetically pleasing compared to workplace images, which is contrary to human perception upon closer examination. This inconsistency raises concerns about the NIMA score's effectiveness.

The age and gender detection model from OpenCV, similarly, exhibited noticeable limitations in its accuracy and trustworthiness. The lower quality of AI-generated images used for testing, as compared to real-life photographs, likely contributed to the model's decreased performance. Out of the two, gender detection demonstrated greater accuracy. However, the age detection model consistently produced inaccurate results, often falsely identifying images without human faces or individuals of Asian descent as babies. These findings highlight the need for further development and refinement.

Overall, these experiment reveals the limitations of current No-reference IQA Metrics and tools in reliably predicting image quality, especially in accurately analyzing semantic quality at an advanced level. These metrics proved to be more effective in evaluating visual quality rather than aesthetic quality. This highlights the need for more advanced and context-aware IQA metrics that can better capture the semantic quality of objects and scenes in images. On a broader level, as generative AI models continue to advance, producing higher-quality images, the focus of evaluation and assessment will need to shift toward more sophisticated quality analysis. Simply striving to attain high scores on existing industry benchmarks such as Fréchet Inception Distance (FID) or Inception Score (IS), will no longer be sufficient. Instead, there is a growing need to develop and refine assessment methodologies that can incorporate, albeit at a generalized estimate level, human-centric evaluations.

On the other hand, the application of the social semiotics framework proved to be rigorous and consistent. The interpretation and ratings of images using this framework were found to be in alignment with the analysis derived from human surveys. However, it is important to note that employing the social semiotics framework required significant time and effort. Therefore, future research should explore ways to streamline the application of this framework without compromising its accuracy.

Implications for future research

The analysis and discussion presented above underscore the significant implications of biases and stereotypes in cultural representation within generative AI image models. These implications call for further research and exploration to address the underlying issues and advance fair and accurate cultural portrayals, including:

Enhanced Bias Detection and Mitigation Techniques: Future research should focus on developing more robust and effective techniques for detecting and mitigating cultural biases within generative AI image models. This entails exploring advanced algorithms, data preprocessing methods, and model architectures that can better identify and redress biases at various stages of model development (Buolamwini & Gebru, 2018).

Intersectionality and Multi-Dimensional Representation: Given the intersectional nature of cultural biases, future research should examine and address biases that intersect with other dimensions of identity, such as race, gender, and socioeconomic status. An intersectional approach is vital to ensuring that cultural representation in AI models is comprehensive, inclusive, and recognizes the complex interplay of various identities (Crenshaw, 1991).

Inclusive Dataset Creation and Curation: The availability of diverse and representative datasets is critical for training unbiased generative AI image models. Future research should concentrate on developing methodologies for creating and curating inclusive datasets that accurately reflect the cultural diversity of different regions, communities, and identities. This includes involving diverse stakeholders in the dataset creation process and considering the power dynamics inherent in dataset curation (Buolamwini & Gebru, 2018). Novel approaches to inclusive dataset creation and curation, as described above, should be further explored and adapted for cultural representation, especially for non-Western colonized cultures.

Cultural Awareness and Contextual Understanding: To ensure culturally sensitive and contextually appropriate image generation, future research should explore techniques that equip AI models with a deeper understanding of cultural nuances, historical context, and local perspectives. This may involve incorporating cultural awareness modules into AI models and drawing from disciplines such as anthropology and cultural studies to enrich their contextual understanding (Said, 1978).

User-Centric Design and Feedback Mechanisms: Engaging end-users and stakeholders in the design and evaluation of generative AI image models is crucial. Future research should focus on integrating user feedback mechanisms and participatory approaches to ensure that models align with the expectations and needs of diverse cultural communities. This can also offer insights into hidden assumptions, and underlying biased regimes, and foster a sense of ownership, agency, and co-creation of culturally-aware AI models (Qadri et al., 2023).

Ethical Considerations and Governance Frameworks: Addressing biases in AI models necessitates robust ethical considerations and governance frameworks that guide the development, deployment, and evaluation of these technologies. Future research should focus on developing frameworks that account for cultural biases, promote transparency, and establish mechanisms for accountability.

Cross-Disciplinary Collaboration: Overall, given the complex, intersectionality nature of biases in cultural representation, interdisciplinary collaboration is essential. Future research should encourage collaboration between researchers and experts from computer science, social sciences, humanities, and cultural studies. This collaboration can provide comprehensive insights, foster dialogue, and generate innovative solutions that account for cultural complexities (Ruha Benjamin, 2021).

By addressing these research areas, future studies can advance the development of inclusive and culturally sensitive generative AI image models. This, in turn, will contribute to fostering greater cultural understanding, reducing stereotypes, and promoting equity in and through AI technologies. These endeavors can push the boundaries of knowledge, promote equitable representation, and contribute to a more just and inclusive society that moves us past the Orientalism and post-colonialism narratives.

Limitations and potential biases

While this study aimed to investigate cultural representation in generative image AI models, specifically focusing on Stable Diffusion, it is important to acknowledge certain limitations and potential biases that may have influenced the findings and interpretations. These limitations should be considered when drawing conclusions and generalizing the results.

Firstly, a notable limitation of this study is the limited generalizability as the analysis of cultural accuracy and fairness was conducted on the specific dataset and prompt selection. The analysis and conclusions may not capture the full range of cultural representation within Stable Diffusion. Future research should strive to expand the dataset, and consider using larger and more diverse datasets that encompass a broader range of cultures, especially those from the global South and consider a broader range of prompts to enhance a more comprehensive understanding of cultural representation accuracy and fairness.

Additionally, the qualitative assessment of cultural representation in Stable Diffusion relied on human evaluation and subjective judgments. While efforts were made to minimize biases and ensure reliability using the triangulation approach, inherent subjectivity, and personal perspectives may still have influenced the assessment. The interpretations and evaluations of cultural accuracy may vary among individuals, and the subjective nature of these judgments should be acknowledged. Future research could explore the integration of a focus group with a group of diverse cultural experts to supplement the social semantics analysis and enhance the assessment of cultural accuracy.

Furthermore, the examination of biases and stereotypes in Stable Diffusion's output, drawing from the theoretical frameworks, mainly cultural studies, Orientalism, and intersectionality, might not have captured the full extent of biases present. While efforts were made to identify and analyze potential biases, the complex nature of cultural biases and stereotypes requires a nuanced and in-depth investigation. The reliance on selected theoretical frameworks and my limited theoretical foundation may have limited the exploration of other potential biases and dimensions of cultural representation. Future studies should consider incorporating additional theoretical frameworks, such as critical race theories, to provide a more comprehensive analysis of biases and stereotypes in AI-generated images.

In conclusion, this study has identified several limitations and potential biases that need to be considered when interpreting the findings. It is crucial for future research to address these limitations by expanding the dataset, considering a broader range of prompts, employing augmented assessment methods, and critically examining the theoretical frameworks used. By doing so, researchers can strive towards a more comprehensive understanding of cultural representation in generative image AI models and work towards mitigating cultural biases and stereotypes for everyone, especially those from non-Western cultures in the global South.

Conclusion

In this study, I have investigated the cultural representation in Stable Diffusion, a generative image AI model. The analysis focused on evaluating cultural accuracy and fairness, as well as identifying potential biases and stereotypes, particularly those affecting non-Western cultures in the global South.

Through our examination of Stable Diffusion's output, I found evidence of both strengths and limitations in terms of cultural representation. While the model demonstrated some degree of cultural accuracy, there were instances where inaccuracies and biases were observed across all cultures. The representation of non-Western cultures from the global South, in particular, revealed the presence of cultural biases and stereotypes, reflecting the persisting influence of Orientalism and post-colonialism.

This study contributes significantly to the field of cultural representation in AI by shedding light on the specific ways in which generative image AI models, such as Stable Diffusion, depict different cultures. By employing a critical lens informed by theories in cultural studies, Orientalism, and intersectionality, we have deepened our understanding of the complexities and nuances involved in cultural representation within AI systems. This research underscores the paramount importance of interrogating biases, promoting cultural diversity, and striving for accurate and fair representation in AI technologies.

The findings of this study have several implications for future research and the development of generative image AI models. Firstly, it highlights the urgent need for further investigations into the development and training of AI models to mitigate biases and stereotypes, particularly when representing non-Western cultures from the global South. It is imperative to recognize cultural diversity as a crucial component of AI fairness that is often overlooked. Therefore, cultural inclusivity should be considered a fundamental pillar throughout all phases of AI development and deployment, involving developers, funders, and regulatory bodies.

Secondly, based on the representation patterns and biases uncovered, this study also suggests a set of mitigation tools to address the identified biases. Through experiments with the CCUB dataset and algorithms, promising results have been observed, indicating the efficacy of fine-tuning pre-trained generative AI models using small, culturally inclusive datasets as an effective strategy to enhance cultural diversity. Moreover, regarding methodology, social semantics is proven to be an effective qualitative tool for cultural representation in AI images. Future research, on the other hand, should explore the integration of user feedback and collaborative approaches, such as focus groups, to augment the assessment of models' cultural fairness. Additionally, the study found limitations in IQA benchmarks' ability to analyze semantic quality and directs to a need for more sophisticated human-centric quality analysis as AI models continue to advance.

While this study provides valuable insights into cultural representation in Stable Diffusion, there are still numerous avenues for further exploration. For instance, examining the impact of cultural biases and stereotypes on social perceptions, user experiences, and the broader societal implications of AI-generated images warrants continued investigation. Additionally, exploring the role of ethics and accountability in AI development and deployment, as well as the potential for co-creating AI models with diverse communities represents an intriguing avenue that has the potential to push the frontiers of AI capability.

In conclusion, this study has underscored the critical importance of critically examining cultural representation in generative image AI models, using a case study on Stable Diffusion. By uncovering the presence of biases and stereotypes, particularly towards non-Western cultures in the global South, it emphasizes the need for continuous improvement and conscious efforts to foster accurate, fair, and inclusive cultural representation in AI technologies, especially for communities with limited economic, technological, and political power. By addressing these challenges head-on, we can unlock the true potential of AI systems as empowering tools that are positive and responsible for all communities. It is our collective responsibility to strive for a future where AI technologies can truly empower people and enable meaningful and equitable representation for all.

Bibliography

ACM. (2018). ACM Code of Ethics and Professional Conduct. Retrieved from https://www.acm.org/code-of-ethics David Pogue. (2023, January 15). Art created by artificial intelligence. https://www.cbsnews.com/news/ai-art-created-by-artificial-intelligence/

Baldwin, D. E., Faulkner, A. F., Hecht, S. A., & Lindsley, R. R. (2005). Redefining Culture: Perspectives Across the Disciplines. Mahwah, NJ: Lawrence Erlbaum Associates.

Bezemer, J., & Jewitt, C. (2009, December 1). Multimodal analysis: : Key issues. https://www.semanticscholar.org/paper/Multimodal-analysis:-:-key-issues-Bezemer-Jewitt/1a69e0409cea20800449808b51a7da3512c3c1b9

Caporale, Vincenzo. “Why Vietnam’s Economic Future Is Bright – and Growing Brighter,” September 29, 2022. https://thediplomat.com/2022/09/why-vietnams-economic-future-is-bright-and-growing-brighter/.

Carminati, L. (2018). Generalizability in Qualitative Research: A Tale of Two Traditions. Qualitative Health Research, 28(13), 2094–2101. https://doi.org/10.1177/1049732318788379 IEEE. (2019). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. IEEE Standards Association."

Chambers, I., & Terranova, T. (2016). Introduction. Inflections of Technoculture: Biodigital Media, Postcolonial Theory and Feminism. Anglistica AION an Interdisciplinary Journal, 18.2, 1–18. https://doi.org/10.19231/angl-aion/201421

Chaofeng Chen. (n.d.). chaofengc/IQA-PyTorch: PyTorch Toolbox for Image Quality Assessment, including LPIPS, FID, NIQE, NRQM(Ma), MUSIQ, NIMA, DBCNN, WaDIQaM, BRISQUE, PI and more... Retrieved July 1, 2023, from https://github.com/chaofengc/IQA-PyTorch/tree/main

Cleanlab. (2023). CleanVision [Python]. Cleanlab. https://github.com/cleanlab/cleanvision (Original work published 2022)

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory, and antiracist politics. University of Chicago Legal Forum, 139-167.

"Deckers, N., Fröbe, M., Kiesel, J., Pandolfo, G., Schröder, C., Stein, B., & Potthast, M. (2023). The Infinite Index: Information Retrieval on Generative Text-To-Image Models. Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, 172–186. https://doi.org/10.1145/3576840.3578327

Dehouche, N., & Dehouche, K. (2023). What’s in a text-to-image prompt? The potential of stable diffusion in visual arts education. Heliyon, 9, e16757. https://doi.org/10.1016/j.heliyon.2023.e16757 Cheney-Lippold, J. (2017). A new algorithmic identity: Soft biopolitics and the modulation of control. Theory, Culture & Society, 34(6), 79-98."

Eltantawy, N. (2013). From Veiling to Blogging: Women and media in the Middle East. Feminist Media Studies, 13(5), 765–769. https://doi.org/10.1080/14680777.2013.838356

Entman, R. M. (1994). Representation and reality in the portrayal of blacks on network television news. Journalism & Mass Communication Quarterly, 71(3), 509-520.

European Commission. (2019). Ethics Guidelines for Trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
Chen, I. Y., Szolovits, P., & Ghassemi, M. (2019). Can AI Help Reduce Disparities in General Medical and Mental Health Care? AMA Journal of Ethics, 21(2), E167-179. https://doi.org/10.1001/amajethics.2019.167"

Febryandini, A. D. (2018, June 1). A Semiotic Analysis of Solo Basahan as a Traditional Javanese Clothes. https://www.semanticscholar.org/paper/A-Semiotic-Analysis-of-Solo-Basahan-as-a-Javanese-Febryandini/ce7d7a9019428e11c42ea686662a9123646ff16e

Feldman, T., & Peake, A. (2021). End-To-End Bias Mitigation: Removing Gender Bias in Deep Learning (arXiv:2104.02532). arXiv. https://doi.org/10.48550/arXiv.2104.02532

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. Journal of Personality and Social Psychology, 82(6), 878-902.

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2019). A comparative study of fairness-enhancing interventions in machine learning. In Proceedings of the conference on fairness, accountability, and transparency (pp. 329-338).

Friedrich, F., Brack, M., Struppek, L., Hintersdorf, D., Schramowski, P., Luccioni, S., & Kersting, K. (2023). Fair Diffusion: Instructing Text-to-Image Generation Models on Fairness (arXiv:2302.10893; Version 2). arXiv. http://arxiv.org/abs/2302.10893

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Polity Press.

Frolov, S., Hinz, T., Raue, F., Hees, J., & Dengel, A. (2021). Adversarial text-to-image synthesis: A review. Neural Networks, 144, 187–209. https://doi.org/10.1016/j.neunet.2021.07.019

Gao, X., Wang, X., Yu, F., & Darrell, T. (2020). Deep bsde solver for high-dimensional partial differential equations. arXiv preprint arXiv:2001.01328.

Gebru, T., et al. (2018). Datasheets for datasets. In Proceedings of the 2018 ACM Conference on Fairness, Accountability, and Transparency, 183-195.

Geirhos, R., et al. (2020). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In International Journal of Computer Vision, 128(4), 1078-1101.

Gietzelt, Dale. “The Indonesianization of West Papua.” *Oceania* 59, no. 3 (1989): 201–21.

Glick, Peter, and Susan T. Fiske. “The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism.” Journal of Personality and Social Psychology 70, no. 3 (March 1996): 491–512. https://doi.org/10.1037/0022-3514.70.3.491.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

Goto-Jones, Christopher. “Magic, Modernity, and Orientalism: Conjuring Representations of Asia.” *Modern Asian Studies* 48, no. 6 (November 2014): 1451–76. https://doi.org/10.1017/S0026749X13000498.

Greenwald, A. G., & Hedges, L. V. (2018). The power of stereotyping and confirmation bias to overwhelm accurate assessment: The case of economics, gender, and risk aversion. American Economic Review, 108(3), 714-747.

Gregorian, Tamar Meguerditchian. “Becoming Malala: A Discourse Analysis of Western and Middle Eastern Print and Broadcast Coverage of Malala Yousafzai from 2012-2017,” December 1, 2018.

Hall, S. (Ed.). (1997). Representation: Cultural representations and signifying practices. Sage in association with the Open University.

Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In Advances in Neural Information Processing Systems, 3315-3323.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models (arXiv:2006.11239). arXiv. https://doi.org/10.48550/arXiv.2006.11239

Hodkinson, J. R., Walker, J., Mazumdar, S., & Feichtinger, J. (2013, December 2). Deploying Orientalism in culture and history: From Germany to Central and Eastern Europe. https://www.semanticscholar.org/paper/Deploying-Orientalism-in-culture-and-history%3A-from-Hodkinson-Walker/2dfe647cd61806024ed7ef680555aba719c4ecb1

Hofstede, G. (1983). Culture’s Consequences: International Differences in Work-Related Values. Administrative Science Quarterly, 28(4), 625. https://doi.org/10.2307/2393017

Holstein, K., Vaughan, J. W., Daumé III, H., Dudík, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–16. https://doi.org/10.1145/3290605.3300830 Carminati, L. (2018). Generalizability in Qualitative Research: A Tale of Two Traditions. Qualitative Health Research, 28(13), 2094–2101. https://doi.org/10.1177/1049732318788379"

hooks, bell. (1992). Black looks: Race and representation. Boston, MA : South End Press. http://archive.org/details/blacklooksracere00hook

Hossein Talebi. (2017). Introducing NIMA: Neural Image Assessment – Google AI Blog. https://ai.googleblog.com/2017/12/introducing-NIMA-neural-image-assessment.html?m=1 Chen, I. Y., Szolovits, P., & Ghassemi, M. (2019). Can AI Help Reduce Disparities in General Medical and Mental Health Care? AMA Journal of Ethics, 21(2), E167-179. https://doi.org/10.1001/amajethics.2019.167"

Hosseini, S. A., et al. (2021). Analyzing and mitigating bias in language translation systems. In Proceedings of the AAAI Conference on Artificial Intelligence, 35(15), 13794-13802.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), Article 9. https://doi.org/10.1038/s42256-019-0088-2

Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. Psychological Review, 100(1), 109-128.

Kamishima, T., Akaho, S., & Asoh, H. (2012). Fairness-aware classifier with prejudice remover regularizer. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 35-46.

Kellner, D., & Share, J. (2007). Critical Media Literacy, Democracy, and the Reconstruction of Education.

Khan, M., & Zahra, S. (2016). Portrayal of Muslim Women by Western Print Media: A Textual Analysis in Feministic & Orientalist Perspective. 17, 41–57.

Kikuchi, K., Inoue, N., Otani, M., Simo-Serra, E., & Yamaguchi, K. (2022). Generative Colorization of Structured Mobile Web Pages. *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3639-3648.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. In Innovations in Theoretical Computer Science, 43-52.

Kronenberg, F. A. (2007). From Trade Cards to the Internet: Depiction of Germans and Germany in American Advertising.

Lee, Y. T., McCauley, C., & Draguns, J. G. (2000). Stereotypes of East Asians and Asians in North America. In Handbook of culture and psychology (pp. 201-222). Oxford University Press.

Leeuwen, G. K., Theo van. (2020). Reading Images: The Grammar of Visual Design (3rd ed.). Routledge. https://doi.org/10.4324/9781003099857

Li, S., Zhang, Y., Zhang, S., Li, J., & Yang, M. (2021). A Stable Diffusion Compression Model for Image and Video Compression. arXiv preprint arXiv:2107.03320.

Liu, F., Bugliarello, E., Ponti, E. M., Reddy, S., Collier, N., & Elliott, D. (2021). Visually Grounded Reasoning across Languages and Cultures (arXiv:2109.13238; Version 1). arXiv. https://doi.org/10.48550/arXiv.2109.13238

Liu, Z., Shin, Y., Okogwu, B.-C., Yun, Y., Coleman, L., Schaldenbrand, P., Kim, J., & Oh, J. (2023). Towards Equitable Representation in Text-to-Image Synthesis Models with the Cross-Cultural Understanding Benchmark (CCUB) Dataset. https://doi.org/10.48550/ARXIV.2301.12073

Long, J., & He, J. (2021). Cultural Semiotics and the Related Interpretation: 2021 International Conference on Public Relations and Social Sciences (ICPRSS 2021), Kunming, China. https://doi.org/10.2991/assehr.k.211020.340

Little, Douglas. *American Orientalism: The United States and the Middle East since 1945*. University of North Carolina Press, 2008. https://www.jstor.org/stable/10.5149/9780807877616_little.

Mallette, Karla. “European Modernity and the Arab Mediterranean: Toward a New Philology and a Counter-Orientalism.” In *European Modernity and the Arab Mediterranean*. University of Pennsylvania Press, 2011. https://doi.org/10.9783/9780812205268.

Matsumoto, K. (2020). Orientalism and the Legacy of Racialized Sexism: Disparate Representational Images of Asian and Eurasian Women in American Culture. Young

Scholars In Writing. https://www.semanticscholar.org/paper/Orientalism-and-the-Legacy-of-Racialized-Sexism%3A-of-Matsumoto/ac611a7a17d8a3552bef5724a325cf8e96d18e3e

McChesney, R. W. (2008). The political economy of media: Enduring issues, emerging dilemmas. New York : Monthly Review Press. http://archive.org/details/politicaleconomy2008mcch

McChesney, R. W. (2008). The political economy of media: Enduring issues, emerging dilemmas. New York : Monthly Review Press. http://archive.org/details/politicaleconomy2008mcchDavid Pogue. (2023, January 15). Art created by artificial intelligence. https://www.cbsnews.com/news/ai-art-created-by-artificial-intelligence/

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2019). The ethics of algorithms: Mapping the debate. Big Data & Society, 6(2), 2053951716679679.

Mladić, D. (2021). Artificial intelligence and global distribution of power. International Studies, 21(2), 113–125.
Mladić, D. (2021). Artificial intelligence and global distribution of power. International Studies, 21(2), 113–125.
Negra, Diane. *The Irish in Us: Irishness, Performativity, and Popular Culture*. Duke University Press, 2006. https://doi.org/10.2307/j.ctv11hpn2b.

Ngueajio, M. K., & Washington, G. (2022). Hey ASR System! Why Aren't You More Inclusive? Automatic Speech Recognition Systems' Bias and Proposed Bias Mitigation Techniques. A Literature Review (Vol. 13518, pp. 421–440). https://doi.org/10.1007/978-3-031-21707-4_30

Nguyen, Hanhthi. "Branding the Others," 2019. https://www.semanticscholar.org/paper/Branding-the-Others-Nguyen/71274339b4cdce234a81a2bf065ef925326a7c11.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

OpenCV: Face analytics pipeline with G-API. (n.d.). Retrieved June 25, 2023, from https://docs.opencv.org/4.x/d8/d24/tutorial_gapi_interactive_face_detection.htmlDeckers, N., Fröbe, M., Kiesel, J., Pandolfo, G., Schröder, C., Stein, B., & Potthast, M. (2023). The Infinite Index: Information Retrieval on Generative Text-To-Image Models. Proceedings of the 2023 Conference on Human Information Interaction and Retrieval, 172–186. https://doi.org/10.1145/3576840.3578327

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. Journal of Personality and Social Psychology, 90(5), 751-783.

Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. In Proceedings of the 34th International Conference on Machine Learning, 1517-1525.

Qadri, R., Shelby, R., Bennett, C. L., & Denton, E. (2023). AI's Regimes of Representation: A Community-centered Study of Text-to-Image Models in South Asia. 2023 ACM Conference on Fairness, Accountability, and Transparency, 506–517. https://doi.org/10.1145/3593013.3594016

Righi, R., Samoili, S., López Cobo, M., Vázquez-Prada Baillet, M., Cardona, M., & De Prato, G. (2020). The AI techno-economic complex System: Worldwide landscape, thematic subdomains and technological collaborations. Telecommunications Policy, 44(6), 101943. https://doi.org/10.1016/j.telpol.2020.101943

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. Personality and Social Psychology Review, 2(1), 40-62.

Ruha Benjamin, A. (2021). Race After Technology: Abolitionist Tools for the New Jim Code. Information, Communication & Society, 24(13), 1989–1990. https://doi.org/10.1080/1369118X.2020.1844269

Sa'di, A. H. (2021). Orientalism in a globalized world: Said in the twenty-first century. Third World Quarterly, 42(11), 2505–2520. https://doi.org/10.1080/01436597.2020.1788935 Said, E. W. (1978). Orientalism (1st ed.). Pantheon Books."

Said, E. W. (1978). Orientalism (1st ed.). Pantheon Books.

Salminen, J., Jung, S., Chowdhury, S., & Jansen, B. J. (2020). Analyzing Demographic Bias in Artificially Generated Facial Pictures. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, 1–8. https://doi.org/10.1145/3334480.3382791

Schwemmer, C., Knight, C., Bello-Pardo, E. D., Oklobdzija, S., Schoonvelde, M., & Lockhart, J. W. (2020). Diagnosing gender bias in image recognition systems. Socius, 6, 2378023120967171. doi:10.1177/2378023120967171

Shome, R. (2019). Thinking Culture and Cultural Studies—From/of the Global South. Communication and Critical/Cultural Studies, 16(3), 196–218. https://doi.org/10.1080/14791420.2019.1648841

Shrestha, S., & Das, S. (2022). Exploring gender biases in ML and AI academic research through systematic literature review. Frontiers in Artificial Intelligence, 5, 976838. https://doi.org/10.3389/frai.2022.976838 Escobar, A. (2018). Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds. Duke University Press."

Sohl-Dickstein, J., Li, Z., & Dinh, L. (2021). A robust self-normalizing diffusion. arXiv preprint arXiv:2105.05233.

Song, X., Ma, C., & Yang, Y. (2020). Improved Techniques for Training Score-Based Generative Models. arXiv preprint arXiv:2006.09011.

Staff, A. F. G., September 17, U., 2021, & Comments51, 11:03 a m Share on Facebook Share on TwitterView. (n.d.). When it comes to the census, Arab Americans are an invisible minority—The Boston Globe. BostonGlobe.Com. Retrieved July 3, 2023, from https://www.bostonglobe.com/2021/09/17/opinion/when-it-comes-census-arab-americans-are-an-invisible-minority/

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797-811.

Stoddart, M. C. J. (2007). Ideology, Hegemony, Discourse: A Critical Review of Theories of Knowledge and Power. Social Thought & Research, 28, 191–225. http://www.jstor.org/stable/23252126

The Global AI Talent Tracker. (2020). MacroPolo. https://macropolo.org/digital-projects/the-global-ai-talent-tracker/

Tschofen, M. (2017). The Representation of East Asian Women in Western Media – Studies in Visual Cultures – ENG 705. https://visualculture.blog.torontomu.ca/the-representation-of-east-asian-women-in-western-media/

Vincent, J. (2023, January 26). BuzzFeed says it will use AI tools from OpenAI to personalize its content—The Verge. https://www.theverge.com/2023/1/26/23572834/buzzfeed-using-ai-tools-personalize-generate-content-openai

Xie, Y., Ma, C., & Zhu, S. (2021). Diffusion probabilistic models for image inpainting. arXiv preprint arXiv:2102.09672.

Vráblíková, Lenka. "Reading the Sexual Economy of Academic Freedom with Sarah Kofman and Jacques Derrida: A Feminist Deconstruction of Kant's Concept of the University." Australian Feminist Studies 35, no. 103 (January 2, 2020): 54–69. https://doi.org/10.1080/08164649.2019.1698285.

Yasin, M. S. M., Hamid, B. A., Othman, Z., Bakar, K. A., Hashim, F., & Mohti, A. (2012). A Visual Analysis of a Malaysian English School Textbook: Gender Matters. Procedia - Social and Behavioral Sciences, 69, 1871–1880. https://doi.org/10.1016/j.sbspro.2012.12.140

You, J., & Korhonen, J. (2022). Attention integrated hierarchical networks for no-reference image quality assessment. Journal of Visual Communication and Image Representation, 82, 103399. https://doi.org/10.1016/j.jvcir.2021.103399

Yu, C. J. (2020). Asian Americans: The Overrepresented Minority?: Dispelling the 'Model Minority' Myth. ASA Monitor, 84(7), 32–33. https://doi.org/10.1097/01.M99.0000688732.01865.74

Żerebecki, B. G., Opree, S. J., Hofhuis, J., & Janssen, S. (2021). Can TV shows promote acceptance of sexual and ethnic minorities? A literature review of television effects on diversity attitudes. Sociology Compass, 15(8), e12906. https://doi.org/10.1111/soc4.12906

Żerebecki, B. G., Opree, S. J., Hofhuis, J., & Janssen, S. (2021). Can TV shows promote acceptance of sexual and ethnic minorities? A literature review of television effects on diversity attitudes. Sociology Compass, 15(8), e12906. https://doi.org/10.1111/soc4.12906

Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335-340.

Zhao, J., et al. (2017). Men also like shopping: Reducing gender bias amplification using corpus-level constraints. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2979-2989.

World Bank Open Data." Accessed July 8, 2023. (Hall, 1997; Stoddart, 2007).

World Bank Open Data." Accessed July 8, 2023. https://data.worldbank.org/country/indonesia.

Appendix

Image Quality Assessment Results

Culture	prompt	Average of BRISQUE_score	SUM of StdDev of BRISQUE_score
American	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	5.89	11.75
	workplace	28.47	7.63
Austrian	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.64	12.30
	workplace	27.72	8.25
Indonesian	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.01	11.56
	workplace	27.72	8.25
Irish	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.56	11.53
	workplace	27.72	8.25
Japanese	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.64	12.30
	workplace	28.47	7.63
Pakistan	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.64	12.30
	workplace	27.72	8.25
Vietnamese	city	10.92	6.55
	clothing	9.90	10.12
	person	3.06	10.57
	woman	6.64	12.30
	workplace	27.72	8.25

Table 3 Cultural Image BRSIQUE score

Culture	*Prompt*	Average of NIMA_score	StdDev of NIMA_score
American	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.28	0.31
	workplace	4.95	0.33
American Total		5.24	0.33
Austrian	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.29	0.32
	workplace	4.91	0.36
Austrian Total		5.23	0.34
Indonesian	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.29	0.30
	workplace	4.91	0.36
Indonesian Total		5.23	0.33
Irish	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.28	0.29
	workplace	4.91	0.36
Irish Total		5.23	0.33
Japanese	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.29	0.32
	workplace	4.95	0.33
Japanese Total		5.24	0.33
Pakistan	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.29	0.32
	workplace	4.91	0.36
Pakistan Total		5.23	0.34
Vietnamese	city	5.33	0.30
	clothing	5.30	0.39
	person	5.34	0.32
	woman	5.29	0.32

	workplace	4.91	0.36
Vietnamese Total		5.23	0.34

Table 4 Cultural Image NIMA score

Survey Results

Culture	AVERAGE of person [Image 1]	AVERAGE of person [Image 2]	AVERAGE of person [Image 3]	AVERAGE of woman [Image 1]	AVERAGE of woman [Image 2]	AVERAGE of woman [Image 3]	AVERAGE of clothing [Image 1]	AVERAGE of clothing [Image 2]	AVERAGE of clothing [Image 3]	AVERAGE of city [Image 1]	AVERAGE of city [Image 2]	AVERAGE of city [Image 3]	AVERAGE of workplace [Image 1]	AVERAGE of workplace [Image 2]	AVERAGE of workplace [Image 3]
American	2.57	3.57	4.14	2.71	3.43	3.43	2.00	3.00	2.71	3.86	3.86	3.71	4.00	4.43	4.00
Austrian	1.57	1.81	2.52	2.62	2.29	1.52	1.43	1.62	1.19	3.62	3.86	3.62	2.86	2.86	2.81
Indonesian	2.09	2.59	1.91	2.34	2.63	2.16	1.66	2.00	1.97	3.09	2.53	2.91	2.69	3.09	3.34
Irish	3.25	2.75	2.25	2.75	2.25	3.50	2.25	2.75	2.25	4.00	4.50	3.50	3.25	2.75	2.75
Japanese	3.50	2.00	3.75	3.00	3.00	3.75	2.75	3.25	4.00	4.00	4.00	4.25	2.50	3.00	4.50
Pakistani	2.40	2.50	2.70	3.30	2.80	3.50	3.70	3.20	1.80	2.20	3.10	3.40	2.70	3.40	3.30
Vietnamese	2.66	2.31	2.44	2.97	2.22	2.34	1.72	1.81	1.91	3.50	4.06	2.38	2.31	3.03	3.06
Grand Total	**2.31**	**2.40**	**2.47**	**2.73**	**2.51**	**2.40**	**1.90**	**2.12**	**1.92**	**3.35**	**3.49**	**3.05**	**2.71**	**3.13**	**3.22**

Table 5 Human rating of best images per prompt per culture

	Did you notice any cultural stereotypes or biases present in the images?		
Culture	**Yes**	**No**	**Grand Total**
American	12	2	14
Austrian	16	5	21
Indonesian	23	9	32
Irish	8	0	8
Japanese	8	1	9
Pakistani	9	1	10
Vietnamese	23	9	32
Grand Total	99	27	126

Table 7 Evaluations of Cultural Bias by Human Evaluators

Survey Open-ended Response

What culture are you identified with? Please select one.	If you noticed any cultural stereotypes or biases, or have additional comments, please elaborate.
Ireland/ Irish	almost every people were red haired on the picture. However not all Irish is red haired. The second picture of the man would be better with a different eye color. I never saw an Irish person with similar icy blue eyes. About the workplace section, usually the offices are in the city center or in an industrial area but it would be nice to have that landscape shown in the pictures.
United States/ American	Images of traditional peoples were very traditional and didn't vary much from generational stereotypes, also inaccurate clothing for native american
Ireland/ Irish	The second set of images was all Asian when I have selected Irish culture
Ireland/ Irish	Redheaded people over-represented. People tended to look like exaggerated versions of Irish people. I would also say that the people were 'uglier' than the average Irish person (Very subjective I know)
Pakistan/ Pakistani	Our white clothing is stereotyped even though many of us wear western clothing too. Pakistanis do not have a standard type of facial structure or skin tone. We have a multitude of that." These stereotypes are based on Punjabi culture.
Indonesia/ Indonesian	"traditional" is made at a generic level
Pakistan/ Pakistani	Consideration of everybody (male) seemingly dressed in a fashion which is pre-colonization era.
Indonesia/ Indonesian	Indonesia is huge country with diverse culture, 33 provinces, we have someone look like African to someone looks like Arab, need to be more spesific which culture that is referred to
Japan/ Japanese	The portrait or image of Japanese people are coming from edo period which nobody wears that kind of makeup and traditional clothes in daily lives. It’s so interesting how our image of people are so heavily influence from this period. Also japan has strong gender stereotypes when it come to clothings so from this aspect i found a bit unusual when man wearing kimono for example. (Doesn’t mean I dislike it, I hope it was beautiful) let me know if you need more feedback or insight! Good luck :))
United States/ American	patriotism
Austria/ Austrian	clothing
Austria/ Austrian	Doesnt represent all of the people. there are many places without Mountains and big churches
United States/ American	The native american images were interesting... like trying to combine the american flag with traditional native american clothing... obviously NOT historically accurate.
Austria/ Austrian	Very traditional clothing
Indonesia/ Indonesian	It is difficult to determine a cultural image of Indonesia because it is so diverse. The images are focused on Borneo/Kalimantan. I think if there is full information about the prompt used to create the image, it will be more transparent and appropriate to judge.

Indonesia/ Indonesian	The faces of the people look the same, the images depicting workplace and city look gentrified
Indonesia/ Indonesian	I especially like the city images. Surely, most of Indonesia is rural, but a few cities like Jakarta have lots of tall buildings. That being said, I feel that there are stereotype biases towards women in the previous page. There are numerous women in Jakarta that don't wear hijabs / blankon.
Pakistan/ Pakistani	The diversity is lost, though similarity in most images can be found in our culture but those will represent only a fraction of the population. Culture (that is based on symbolism, festivals, fashion) differs significantly across geographical regions and social class. Moreover, culture can be a lot more than what can be observed through pictures, for example, the values and view on morality, psyche and approach to a crisis. Limiting the complex culture to a few images representing just the 'one way' overwhelmingly leads to stereotypes.
Pakistan/ Pakistani	Not every pakistani woman covers her head, not every pakistani man wears shalwar kameez especially in a workplace and all the images look like poor people
Indonesia/ Indonesian	More likely it is Vietnamese or cambodian
Austria/ Austrian	It is very problematic and not correct to equal culture with nationality and speak of "cultures" in plural. Culture is not separate bubbles bound to national borders. Although this unreflecting terminology is often used in everyday conversations, it must not be used in a scientific paper rashly. National stereotypes based on traditional clothes and historical photographs etc. might be part of collective culture but do not represent the everyday understanding, functions and logics of society.
Indonesia/ Indonesian	asians are racially identified as chinese only
Pakistan/ Pakistani	I have an opinion regarding the term "Pakistani Culture". Pakistan is historically diverse country with the history of more than 5000 years (Indus velly) and cultures like Punjabi, Sindhi, Baloch, Pashtun, Kashmiri, Siraiki, Makrani, Balti etc. So we must take all these accounts into detail while talking about "Pakistani" Culture.
Indonesia/ Indonesian	Many cultures exist in Indonesia makes it difficult to determine if an individual falls inline of our expectations of what an Indonesian is supposed to look like or not. Some images may display very exotic individuals who do not appear Indonesian according to an observer's biases despite the individual being a fully native Indonesian.
Pakistan/ Pakistani	Almost all photos only represent the historical/traditional scene and not the modern picture you would see today, especially in cities.
Indonesia/ Indonesian	Those symbols on the fabric doesnt always represent our culture
Indonesia/ Indonesian	Color, size, ethnic symbolism, lack of diverse representation
Indonesia/ Indonesian	There is a bias of the face shapes. also the skin color is too dark or red
Vietnam/ Vietnamese	The problem is that what is meant by "your culture"? Vietnamese culture is not homogeneous. It is so diverse that I doubt if you can collect any good data from such a questionnaire.
Vietnam/ Vietnamese	Culture is something quite hard to express verbally, you have to feel it.

Vietnam/ Vietnamese	Asian = small eyes; lack of variety in cultural attire; lack of nuance to city/workplace (too generic)
Indonesia/ Indonesian	Focus of features is more on indigenous than the general look of an everyday people
Vietnam/ Vietnamese	Vietnamese people have dark skin colour; (possibly due to this research prompts to AI generator) There are only men and women, middle aged, non-disabled people that appear in this survey. But in the end, I don't really understand what it means by 'cultural representation' in this study. Culture should be more complex than aspects of human, landscape, material and design.
Pakistan/ Pakistani	It represents Indian culture. Pakistani men don’t wear jewellery. And the turban style of men was Sikh religious style. Only the elder villagers wear turban and the style is very casual and simple.
Vietnam/ Vietnamese	Skin tone of men is dark, of women is bright which is not so true. Most clothes are traditional clothes which are not usually worn in daily life. The workplace and street photos are more realistics.
Vietnam/ Vietnamese	That Vietnamese people are first seen poor and live a more agricultural life, and that we usually wear our traditional cone hat, which is wrongly depicted by AI. The "traditional dresses" in these pictures are heavily Chinese influenced, which I personally disagree.
Vietnam/ Vietnamese	The people wearing conical hats and the small eyes, high cheek bones
Vietnam/ Vietnamese	Focus too much on old generation, thế poor and village
Vietnam/ Vietnamese	Some typical face characteristic
Vietnam/ Vietnamese	Lots of bias and similarities to Chinese culture
Indonesia/ Indonesian	Indonesia is a tropical nation with lots of body of water, and more oft than not beaches are featured. But we have no venice-like cities where our streets are made of water
Pakistan/ Pakistani	most of the clothing were based on Islamic culture rather than Persian/Iranian
Indonesia/ Indonesian	matanya banyak yg sipit kek bukan org indo
Austria/ Austrian	Mountains, churches. Traditional garbs of a certain style I've never seen in Austria tho
Ireland/ Irish	Broken nose implying fighting. Photos of woman looked like medieval peasants
Austria/ Austrian	Traditional Clothing looks old
Indonesia/ Indonesian	They're all brown, and the traditional clothing only represent ones from certain places
United States/ American	Lots of American flags
Indonesia/ Indonesian	Most of them have southeast asia main land face and clothing feature such as thais, laos, cambodia. Its like showing french picture and ask whether they look like germany
Indonesia/ Indonesian	mosty moslem
Indonesia/ Indonesian	that we look more East Asian, that our cities are submerged/are built in large bodies of water

Pakistan/ Pakistani	These are mostly men. Also only Muslims with Arab look. This is like depicting Americans as Mexicans because they are close geographically and they share the same religion.
Indonesia/ Indonesian	Its not even "Indonesian"
Indonesia/ Indonesian	Most of the images depicts some level of east Asian clothes, facial features, and culture. These images not even remotely close to south east asian, let alone Indonesian
Indonesia/ Indonesian	The people pictures look more East Asian than Indonesian.
Japan/ Japanese	yes, man is wearing women looking clothes

Table 8 Opened ended response regarding Cultural Bias Across Cultures

Survey Form

How does Generative AI see you? Exploring cultural representations in Stable Diffusion Images

Introduction

Dear participant,

Please fill out this short survey to evaluate images of different cultures generated by Stable Diffusion. Your feedback will help us understand the cultural awareness of the model and help build a more accurate and equitable AI for your culture!

Quick overview:

All images are generated with Stable Diffusion v1.5, a text-to-image generative AI model, using the same setup, i.e., the same prompts and parameters. E.g., a photo of a (American) person, realistic.

The images selected in this survey are the ones with the highest aesthetic score (NIMA) and then manually reviewed and chosen for their appropriateness and representativeness of the respective dataset.

Please answer the following questions based on your impressions of the images. Your responses will be kept confidential and will only be used for research purposes. The result will be included in my thesis project titled "Cultural Representation in Generative AI Image Model: A cross-disciplinary case study of Stable Diffusion". If you have any question or feedback, please feel free to reach out to me with the email provided at the end of this survey.

Survey section

1. **Cultural identity**

What culture are you identified with?
Please select one. If you are identified with 2 or more, please choose the one you feel more strongly identified with.

[Drop down list of 7 options of 7 cultures to select]

2. **Cultural accuracy assessment**

Please see the images generated about your culture and answer the questions below.

Ratings are arranged in row order, from left to right, e.g., Image 1 = Image in the Left corner.

Rating is on a scale from 1 -5, with 1 = not at all appropriate, 5 = very appropriate.

[Insert 3 best images of each prompt that correspond to the respondent’s culture. Below is a sample of the question for Images of Austrian people]

3. **Cultural stereotype presence**

Did you notice any cultural stereotypes or biases present in the images?

- Yes
- No

4. **Open-ended question**

If you noticed any cultural stereotypes or biases, or have additional comments, please elaborate.

CCUB and Stable Diffusion Comparison

Mode	Prompt: "photo of Japanese person in traditional clothing, realistic"				
Stable Diffusion v1.5					
CCUB					

Table 9 CCUB and Stable Diffusion Comparison